Z. C.

MIHAI ŢARCĂ

STATISTICĂ

II

UNIVERSITATEA „AL. I. CUZA“ IAȘI

Scanned with OKEN Scanner

MIHAI ȚARCĂ

STATISTICĂ

II

UNIVERSITATEA „AL. I. CUZA" IAȘI
FACULTATEA DE ȘTIINȚE ECONOMICE
1979

Scanned with OKEN Scanner

S T R U C T U R A C U R S U L U I

Volumul I

Secţiunea I. INFORMATIA STATISTICA

Cap.I. Problemele generale ale statisticii ca ştiinţă
Cap.II. Noţiuni fundamentale ale statisticii
Cap.III. Obţinerea informaţiilor statistice
Cap.IV. Gruparea şi centralizarea informaţiilor statistice
Cap.V. Prezentarea informaţiilor statistice

Secţiunea II. ANALIZA SERIILOR DE REPARTITIE

Cap.VI. Mărimile medii
Cap.VII. Indicatorii variaţiei
Cap.VIII. Repartiţii ale unor variabile aleatoare
Cap.IX. Caracteristici ale formei distribuţiilor

Secţiunea III. SONDAJE STATISTICE

Cap.X. Legea numerelor mari
Cap.XI. Sondaje statistice
Cap.XII. Verificarea ipotezelor statistice

Volumul II

Secţiunea IV. ANALIZA DISTRIBUTIILOR MULTIDIMENSIONALE

Cap.XIII. Regresie şi corelaţie
Cap.XIV. Regresie şi corelaţie simplă liniară
Cap.XV. Regresie şi corelaţie simplă curbilinie
Cap.XVI. Regresie şi corelaţie multiplă
Cap.XVII. Corelaţia neparametrică
Cap.XVIII. Asociaţia între caracteristici calitative.

Secţiunea V. ANALIZA SERIILOR DINAMICE

Cap.XIX. Indicatorii seriilor dinamice
Cap.XX. Tendinţa generală de evoluţie a fenomenelor în timp
Cap.XXI. Indicii

BIBLIOGRAFIE GENERALA SELECTIVA

Scanned with OKEN Scanner

C U P R I N S

pag.

13. REGRESIE SI CORELATIE 1
13.1. Noţiuni preliminare 1
13.2. Metode elementare de studiere a legăturilor dintre fenomene 5
13.2.1. Metoda seriilor statistice interdependente. 5
13.2.2. Metoda grupărilor statistice. 6
13.2.3. Metoda grafică .. . .

14. REGRESIE SI CORELATIE SIMPLA LINIARA 11
14.1. Regresie simplă liniară 11
14.1.1. Metode de rezolvare a sistemului de ecuaţii normale liniare 15
14.1.2. Dualitatea în regresie. 25
14.2. Intensitatea legăturii liniare. 37
14.2.1. Coeficientul de corelaţie 37
14.2.2. Raportul de corelaţie 48
14.3. Legătura dintre coeficienţii de regresie şi coeficientul de corelaţie 58

15. REGRESIE SI CORELATIE SIMPLA CURBILINIE 61
15.1. Modele de regresie simplă curbilinie. 61
15.1.1. Legături de tip parabolic 62
15.1.2. Legături de tip hiperbolic. 88
15.1.3. Legături după modelul unei funcţii putere. 106
15.1.4. Legături de tip exponenţial 114
15.1.5. Legături de tip logaritmic. 120
15.2. Alegerea celei mai potrivite ecuaţii de regresie. . . 124

16. REGRESIE SI CORELATIE MULTIPLA. 127
16.1. Noţiunea de legătură multiplă 127
16.2. Regresie şi corelaţie liniară multiplă. 129
16.2.1. Regresie liniară multiplă 129
16.2.2. Intensitatea corelaţiei multiple liniare. 139

Pag.

16.2.3. Corelaţie parţială 148
16.2.4. Determinaţie multiplă şi parţială. 150
16.3. Regresie şi corelaţie multiplă curbilinie. 157
16.3.1. Regresie multiplă curbilinie 157
16.3.2. Intensitatea corelaţiei multiple curbilinii. . . . 161
16.3.3. Determinaţia curbilinie multiplă şi parţială . . . 163
16.3.4. Liniarizarea legăturilor multiple curbilinii . . . 165

17. CORELATIA NEPARAMETRICA 167

17.1. Coeficientul de corelaţie a rangurilor al lui Spearman 168
17.2. Coeficientul de corelaţie a rangurilor al lui Kendall 172
17.3. Coeficientul de concordanţă al lui Fechner.. 175

18. ASOCIATIA INTRE CARACTERISTICI CALITATIVE 179

18.1. Noţiunea de asociaţie 179
18.2. Existenţa asociaţiei 180
18.3. Măsurarea intensităţii asociaţiei. 188

19. INDICATORII SERIILOR DINAMICE 193

19.1. Indicatorii nivelului şi sporului absolut. 194
19.2. Ritmul variaţiei totale a fenomenelor. 201
19.3. Ritmul creşterii fenomenelor 203
19.3.1. Ritmul mediu de creştere. 205
19.3.1.1. Metoda mediei geometrice 206
19.3.1.2. Metoda mediei parabolice. 208
19.3.1.3. Metoda trendului exponenţial 209
19.3.2. Determinarea ritmului mediu pentru fenomene complexe 213
19.4. Prognoza fenomenelor pe baza indicatorilor mediі de creştere 219
19.4.1. Metode de prognoză bazate pe sporul mediu. 220
19.4.2. Metode de prognoză bazate pe ritmul mediu. 225

20. TENDINŢA GENERALA DE EVOLUTIE A FENOMENELOR IN TIMP. . 231
20.1. Noţiunea de tendinţă generală de dezvoltare. 231
20.2. Metode de ajustare a seriilor dinamice. 232
20.2.1. Metoda grafică 232

pag.

20.2.2. Metoda mediilor mobile 234
20.2.3. Metoda mediilor înjumătăţite 237
20.2.4. Metoda sporului mediu. 239
20.2.5. Metoda ritmului mediu. 242
20.2.6. Metoda celor mai mici pătrate. 244
20.2.6.1. Tendinţa liniară 245
20.2.6.2. Tendinţa după o parabolă de gradul doi 251
20.2.6.3. Tendinţa după o parabolă de gradul trei. 257
20.2.6.4. Tendinţa exponenţială. 260
20.2.6.5. Ajustarea după o curbă exponenţială modificată 264
20.2.6.6. Curba lui Gompertz 273
20.2.6.7. Curba logistică 279
20.2.6.8. Curba exponenţială de gradul doi 288

21. INDICII 291
21.1. Noţiunea şi importanţa indicilor. 291
21.2. Clasificarea indicilor 293
21.3. Indicii dinamicii 294
21.3.1. Indici simpli sau elementari 294
21.3.2. Indici de grup sau sintetici 299
21.3.2.1. Indici agregaţi 299
21.3.2.2. Indicele mediu aritmetic ponderat 315
21.3.2.3. Indicele mediu armonic ponderat 316
21.3.2.4. Indicele mediu geometric 317
21.3.2.5. Indici calculaţi ca raport între două medii . . . 320
21.4. Erorile de măsurare în construcţia indicilor 327

BIBLIOGRAFIE GENERALA SELECTIVA 329

13. REGRESIE SI CORELATIE

13.1. Noţiuni preliminare

Statistica - formă fundamentală a cunoaşterii ştiinţifice - folosindu-se de metodele ei de cercetare contribuie la descoperirea legilor, legităţilor şi regularităţilor care se manifestă în interiorul şi între fenomenele colective. De altfel, problema fundamentală a oricărei cunoaşteri ştiinţifice, a oricărei ştiinţe este descoperirea legilor care se manifestă între fenomenele care fac obiectul lor de studiu.

Fenomenele şi procesele din natură şi societate sînt rezultatul acţiunii unui mare număr de factori principali şi secundari, esenţiali şi neesenţiali, cuantificabili şi necuantificabili sau cuantificabili cu aproximaţie.

In capitolele precedente ale acestui curs, au fost prezentate deja unele metode şi tehnici susceptibile de a facilita comparaţia între două sau mai multe fenomene colective cum ar fi: privirea comparativă a două serii paralele; compararea mediilor lor şi a gradului de împrăştiere, reprezentarea grafică, tabelul cu dublă intrare etc.

In acest capitol vom depăşi această primă fază descriptivă de studiere a legăturilor dintre fenomene, introducînd elemente noi de măsurare mai precisă a legăturilor dintre ele.

Fenomenele naturale, care iau naştere ca urmare a mişcării mecanice, fizice, chimice, şi biologice a materiei, sînt legate între ele prin legături funcţionale.

De exemplu, fiind dată o bară metalică supusă unor temperaturi diferite: $t_1, t_2, t_3, \dots, t_n$ ea se va dilata şi va lua lungimea, respectiv: $l_1, l_2, l_3, \dots, l_n$. De fiecare dată cînd bara va fi supusă temperaturii t_i va lua lungimea l_i. Astfel putem spune că la o temperatură (dintr-un anumit interval al temperaturilor) îi va corespunde (sau i se poate asocia) o anumită lungime a barei şi numai una. Legătura dintre temperatură şi lungimea barei este o legătură funcţională. Acest tip de legături nu este aplicabil fenomenelor sociale şi economice.

Fenomenelor social-economice le este caracteristic faptul că asupra lor acţionează, alături de cauze determinante pentru legătura sau dependenţa cercetată, numeroşi alţi factori accidentali a căror acţiune poate influenţa diferite fenomene

cercetate. Acest fapt determină existenţa unor legături foarte variate între fenomene. Natura legăturii se stabileşte pe baza unei analize calitative multilaterale. In unele cazuri între fenomene există o <u>legătură cauzală</u> cînd unul dintre fenomene este cauză iar celălalt efect. De exemplu, creşterea productivităţii muncii determină reducerea costului producţiei. Intre alte fenomene poate exista o <u>legătură reciprocă</u>, o interacţiune reciprocă, cînd nici unul dintre fenomene nu poate fi considerat cauză sau efect.

Astfel se constată, cu caracter de legitate, că vîrsta la căsătorie a soţului concordă cu cea a soţiei deci se căsătoresc în general persoane care au o vîrstă apropiată. In acest caz nici vîrsta soţului şi nici cea a soţiei nu poate fi considerată cauză sau efect.

In alte cazuri variaţia paralelă a două fenomene nu <u>se datoreşte legăturii dintre ele, ci influenţei unui al treilea factor</u>, care le determină pe ambele. Aşa de exemplu, cererea de mărfuri şi economiile băneşti ale populaţiei variază în acelaşi sens. Deşi aceste două fenomene variază în acelaşi timp şi în acelaşi sens, nici unul dintre ele nu poate fi considerat cauză sau efect al celuilalt. Ambele fenomene sînt influenţate de un al treilea factor şi anume de veniturile băneşti ale populaţiei.

Din această cauză în cadrul fenomenelor sociale şi economice legătura nu se manifestă în fiecare caz în parte - aşa cum se întîmplă în natură - ci în general şi în medie, - într-un mare număr de cazuri. Fenomenul efect este rezultatul conjugării influenţei mai multor fenomene - cauză şi în acest sistem de legături nu toate raporturile de dependenţă au aceeaşi importanţă, iar acţiunea unora dintre ele se compensează reciproc. Studierea acestor fenomene, cunoaşterea legilor, legităţilor şi regularităţilor de producere a lor nu se poate face pe cale experimentală, ci folosind în principal metoda abstractizării, prin care se înlătură ceea ce este întîmplător şi necesenţial în dezvoltarea lor, desprinzînd astfel raporturile obiective de cauzalitate şi descoperind condiţiile de formare şi dezvoltare ale acestora.

Variaţiile cantitative ale caracteristicii efect (Y) pot fi mai mari sau mai mici decît variaţiile cantitative determinate de caracteristicile cauzale , (X_i) explicite. Unele valori ale caracteristicii efect (Y) pot avea chiar sens contrar sensului de variaţie pe care-l au caracteristicile cauzale (X_i). Cu alte cuvinte între caracteristicile fenomenelor şi proceselor sociale colective există o dependenţă, o <u>legătură stochastică</u>.

Legăturile statistice sau stochastice se caracterizează prin faptul că o caracteristică (X) exercită o anumită influenţă asupra unei caracteristici (Y), îi determină într-o anumită măsură variaţia, însă asupra caracteristicii (Y) mai acţionează şi alte caracteristici în afară de (X), care din punct de vedere al legăturii X şi Y se consideră întîmplătoare.

De exemplu între înzestrarea tehnică a muncii şi nivelul productivităţii muncii există o legătură bine determinată în sensul că pe măsura creşterii nivelului înzestrării tehnice a muncii creşte şi nivelul productivităţii muncii; dar nivelul productivităţii muncii este influenţat şi de alţi factori în afară de înzestrarea tehnică, cum ar fi: experienţa în producţie, nivelul de calificare, retribuirea muncii etc.

Legătura statistică dintre două sau mai multe caracteristici exprimate numeric poartă denumirea de corelaţie. In general, se poate spune că între două sau multe fenomene există corelaţie dacă între ele există o anumită legătură logică.

In opoziţie cu legăturile funcţionale se pot întîlni în realitate şi fenomene între care să nu existe nici un fel de legătură, adică fenomenele respective sînt independente între ele. Deci corelaţia este intermediară între independenţa absolută dintre fenomene şi legăturile funcţionale.

Studiul corelaţiei statistice este îndreptat spre stabilirea existenţei unei legături reale între fenomenele cercetate şi apoi, cu ajutorul unor metode şi tehnici adecvate se caracterizează direcţia, forma şi intensitatea legăturii.

Mulţimea legăturilor statistice se poate clasifica după criterii diferite cum ar fi: numărul caracteristicilor corelate, modul de exprimare a caracteristicii, direcţia legăturii, forma legăturii şi intensitatea ei.

1) După numărul caracteristicilor corelate, legăturile dintre fenomene se pot împărţi în două grupe: legături simple şi legături multiple. Legăturile simple se stabilesc numai între două caracteristici, în timp ce legăturile dintre trei sau mai multe caracteristici poartă denumirea de legături multiple. O astfel de legătură există între recolta medie la hectar, pe de o parte şi calitatea solului, măsurile agrotehnice, condiţiile meteorologice, pe de altă parte. Un alt exemplu: legătura dintre

producţia globală indutrială pe judeţe, fondurile fixe şi numărul de personal din industrie etc.

Dacă legăturile dintre fenomene sînt cauzale, atunci caracteristicicle care joacă rol de cauză poartă denumirea de variabile sau caracteristici independente, factoriale sau determinante şi se notează cu X_1, X_2, ..., X_n , iar caracteristica efect se mai numeşte şi caracteristică sau variabilă determinată sa variablă dependentă sau variabilă rezultativă şi se notează cu Y.

2) După direcţia lor, legăturile statistice pot fi directe sau pozitive şi inverse sau negative. Legăturile directe apar atunci cînd atît variabila (sau variabiele) factorială cît şi variabila rezultativă variază mereu în acelaşi sens. Dacă variaţia caracteristicilor factoriale are sens contrar variaţiei caracteristicii rezultative atunci legătura dintre ele este inversă.

De exemplu, între înzestrarea tehnică a muncii şi nivelul productivităţii muncii există o legătură directă, iar între productivitatea muncii şi costul producţiei există o legătură inversă.

3) După forma lor, legăturile dintre variabile pot fi liniare şi curbilinii (hiperbolă, parabolă, exponenţială etc.).

Orice studiu statistic concret de corelaţie trebuie să pornească de la identificarea existenţei unei legături reale între fenomene şi apoi să treacă la măsurarea statistică a ei. De asemenea, trebuie să se ţină seama de faptul că orice studiu statistic concret se sprijină pe un număr limitat de observaţii şi că, pe baza rezultatelor obţinute prin prelucrarea şi analiza multilaterală a acestora, se încearcă apoi caracterizarea colectivităţii generale. De aceea, rezultatele calculelor de regresie şi corelaţie trebuie să fie interpretate probabilistic.

Pentru studierea legăturilor statistice dintre fenomene se pot utiliza în practică metode elementare (metoda seriilor paralele interdependente, metoda grupărilor, metoda tabelului de corelaţie, metoda balanţelor, metoda grafică etc.) şi metode analitice (regresie şi corelaţie simplă şi multiplă, determinaţia simplă şi multiplă, verificarea ipotezei existenţei legăturii, verificarea semnificaţiei indicatorilor de

regresie, corelaţie şi determinaţie).

Metodele statistice elementare de studiere a dependenţelor dintre fenomene sînt uşor de aplicat, se bazează pe analiza calitativă a variabilelor corelate şi oferă informaţii importante asupra naturii şi caracteristicilor legăturii cercetate. Ele servesc ca punct de plecare în vederea aplicării metodelor analitice mai complexe şi mai precise.

13.2. Metode elementare de studiere a legăturilor dintre fenomene

13.2.1. Metoda seriilor statistice interdependente

Această metodă constă în stabilirea legăturilor dintre fenomene pe baza comparării indicatorilor proveniţi din două serii paralele, una care prezintă variaţia variabilei factoriale X şi alta pentru variabila rezultativă Y. Se pot supune comparaţiei două serii de timp, două serii de spaţiu sau două serii de distribuţie. In primul caz, termenii celor două serii dinamice se înscriu în ordine cronologică, iar în cazul al doilea şi al treilea se procedează la ordonarea termenilor seriilor în ordine crescătoare sau descrescătoare a variabilei factoriale X şi în paralel se înscriu termenii corespunzători ai variabilei rezultative Y. Prin aplicarea acestei metode este posibil să se stabilească existenţa şi direcţia legăturii dintre variabile, dar nu se poate aprecia forma şi intensitatea ei.

In cazul seriilor cronologice interdependente este necesar ca, pe lîngă identificarea legăturilor dintre fenomene să se precizeze şi perioada pentru care are loc legătura respectivă. Astfel se pot evidenţia <u>legături sincrone</u> dintre fenomene, cînd variaţia variabilei factoriale determină concomitent variaţia variabilei rezultative şi <u>legături asincrone</u> în cazul în care variaţia variabilei factoriale se exercită asupra variabilei rezultative după o anumită perioadă de timp, deci cu un anumit decalaj de timp.

Procedeul seriilor paralele interdependente se ppate aplica atît în cazul legăturilor simple cît şi în cazul legăturilor multiple. In cazul legăturilor multiple de corelaţie este necesar ca pe baza analizei critice calitative să se procedeze la ierarhizarea influenţei factorilor cauzali în funcţie

de mărimea influenţei lor asupra variabilei rezultative. Dacă variabilele factoriale nu sînt independente între ele, în cazul corelaţiei multiple, apare fenomenul de multicoliniaritate care trebuie diminuat cît mai mult posibil înainte de a se trece la calculul indicatorilor de regresie şi corelaţie multiplă. Există în literatura de specialitate metode de verificare a existenţei multicoliniarităţii şi de diminuare a efectelor ei în cazul corelaţiei liniare multiple, eliminarea ei totală nu este posibilă.

Aplicarea metodei seriilor paralele interdependente devine mai dificilă în cazul seriilor formate dintr-un număr foarte mare de termeni şi implică în mare măsură subiectivismul cercetătorului. Pentru a se înlătura acest neajuns al metodei seriilor paralele interdependente se utilizează în statistică metoda grupărilor statistice.

13.2.2. Metoda grupărilor statistice

Metoda grupărilor reprezintă o metodă de bază în analiza statistică a legăturilor dintre fenomene şi constă în împărţirea unităţilor statistice după variaţia variabilei X şi calcularea indicatorilor sintetici (mărimi medii, dispersii, mărimi relative de structură) ai caracteristicii rezultative Y pentru fiecare grupă în care a fost împărţită colectivitatea. Această metodă oferă posibilitatea aprecierii existenţei legăturii şi a sensului ei, dar ca şi în cazul metodei anterioare, aprecierea formei şi intensităţii legăturii este cu totul subiectivă, depinzînd în mare măsură de competenţa şi experienţa cercetătorului.

In cazul legăturilor simple se utilizează grupările simple, iar în cazul legăturilor multiple se vor utiliza grupările combinate.

O formă mai desvoltată a acestei metode este metoda tabelului de corelaţie care se utilizează în practică atunci cînd numărul perechilor de valori ale variabilelor corelate este foarte mare. In cadrul tabelului de corelaţie gruparea se poate face pe variantele celor două variabile sau pe intervale de variaţie cînd numărul variantelor este mare. La stabilirea intervalelor de variaţie ale celor două variabile este necesar să se considere, pentru fiecare variabilă, intervale egale de variaţie şi închise.

In cadrul tabelului de corelaţie gruparea se face concomitent după ambele variabile X şi Y. Se recomandă ca numărul intervalelor de variaţie să fie acelaşi pentru ambele variabile.

Scanned with OKEN Scanner

Cu ajutorul tabelului de corelaţie se poate stabili existenţa legăturii dintre variabile, direcţia şi intensitatea ei. Astfel, dacă frecvenţele n_{ij} se dispersează relativ uniform în întregul spaţiu al tabelului, înseamnă că între variabile nu există nici un fel de legătură. Dimpotrivă, dacă frecvenţele se concentrează în jurul unei drepte sau a unei curbe de un anumit tip, înseamnă că între varibile există o legătură a cărei formă este dată de dreapta sau curba respectivă. Cu cît concentrarea frecvenţelor în jurul dreptei sau curbei este mai mare, cu atît se poate aprecia că intensitatea legăturii dintre variabile este mai mare şi invers. Dacă aliura curbei sau dreptei se îndreaptă dinspre valorile mici ale celor două variabile către valorile lor mai mari, atunci se poate spune că între variabile există o legătură directă şi invers. Această presupunere este mai aproape de realitate în cazul legăturilor de formă liniară. O expresie mai precisă a formei legăturilor dintre variabile se va obţine prin reprezentarea grafică a punctelor de coordonate x_i şi mediile condiţionate ale variabilei rezultative $\bar{y}/x_i$.

O formă particulară a metodei tabelului de corelaţie este <u>metoda balanţelor</u>.

13.2.3. Metoda grafică

Această metodă ne permite prezentarea într-o formă geometrică, expresivă şi vizuală a legăturii dintre două variabile.

Metoda grafică se bazează pe sistemul de axe rectangulare în care valorile variabilei factoriale se trec pe abscisă, iar cele ale variabilei rezultative pe ordonată. Perechile de valori corelate x_i, y_i se reprezintă prin cîte un punct în cadrul graficului.

Reprezentarea acestpr puncte în grafic oferă imaginea cîmpului de corelaţie asemănătoare cu un "nor de puncte" , de aici derivă şi denumirea acestui procedeu.

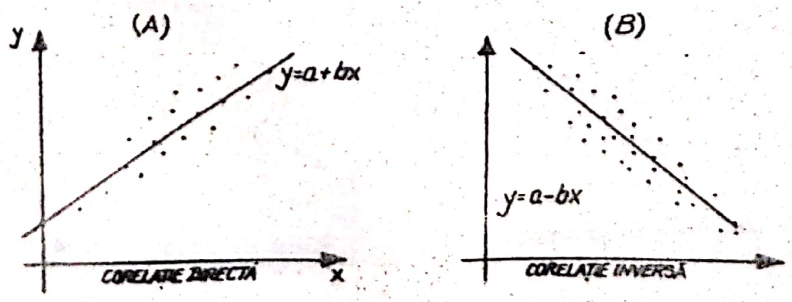


Fig.13. FORME ALE NORULUI DE PUNCTE PENTRU DOUĂ VARIABILE CORELATE LINIAR

După modul de distribuire al acestor puncte în spaţiul graficului se poate aprecia existenţa, forma, direcţia şi intensitatea legăturii dintre două variabile. Concentrarea punctelor în jurul anumitor drepte sau curbe ne va indica faptul că între variabile există o legătură (fig. 13.1 (A) şi (B)), iar dreapta sau curba respectivă va indica forma legăturii dintre variabile. Dacă punctele empirice se dispersează aproape uniform pe între spaţiul graficului, aceasta înseamnă că între variabilele cercetate nu există nici un fel de legătură (fig. 13.I (A)). De asemenea, se poate aprecia că nu există legătură între variabile dacă punctele empirice se concentrează în jurul unei drepte care este paralelă cu una dintre axe (fig. 13.2 (B) şi (C)).

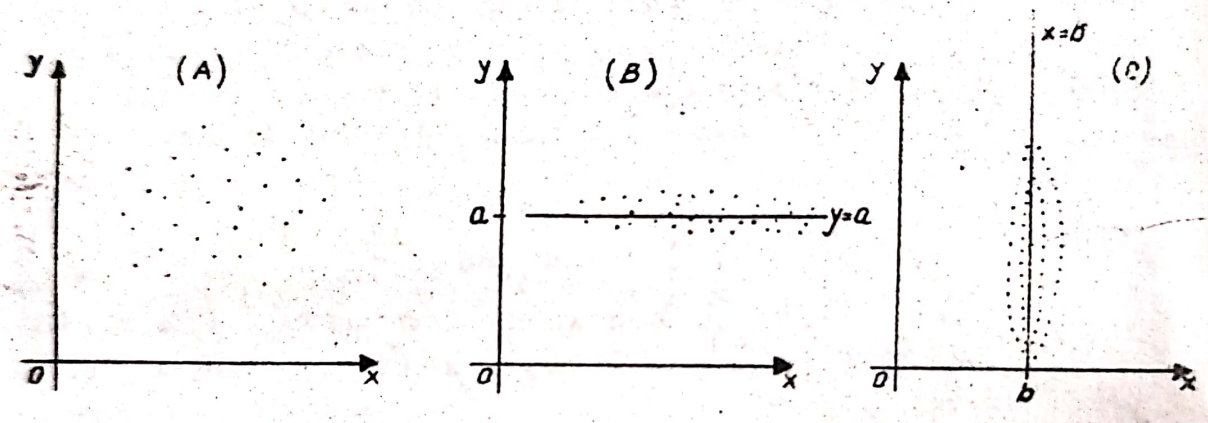


Fig. 13.2. FORME ALE NORULUI DE PUNCTE ÎN CAZUL A DOUĂ VARIABILE INDEPENDENTE

Dacă punctele empirice se situează chiar pe o dreaptă (fig. 13.3 (A) şi (B)) sau pe o curbă de un anumit tip, atunci legătura dintre variabile este o legătură de tip funcţional.

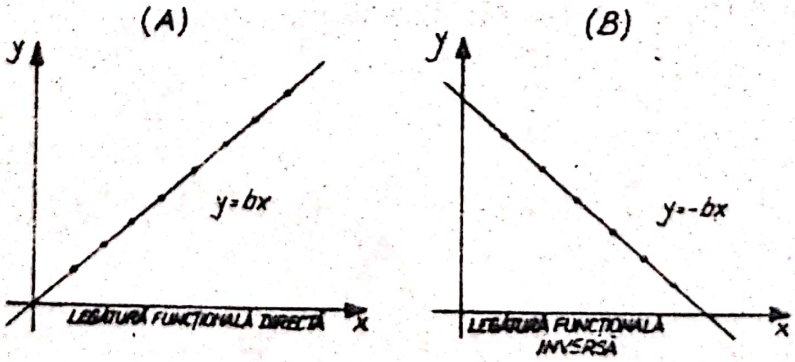


Fig. 13.3. LEGĂTURI FUNCŢIONALE LINIARE ÎN CAZUL A DOUĂ VARIABILE

In funcţie de modul de orientare în planul xoy, a dreptei sau curbei, ce descrie legătura dintre variabile, se poate aprecia direcţia legăturii dintre ele.

Dacă variabilele corelate se prezintă sub forma unui tabel de corelaţie, atunci pentru reprezentarea grafică se vor utiliza procedeele cunoscute: norul de puncte, stereograma, diagrama prin paralelipipede şi suprafaţa poliedrală (vezi cap. 5).

Scanned with OKEN Scanner

14. REGRESIE SI CORELATIE SIMPLA LINIARA

14.1. Regresie simplă liniară

Cercetarea concretă a legăturii dintre fenomene numai pe baza metodelor elementare este insuficientă. Informaţiile obţinute cu acest prilej trebuie completate şi determinate cu precizie mai mare utilizînd metodele analitice. Cu ajutorul acestora se pot calcula indicatori sintetici privind expresia matematică a legăturii dintre fenomene, a intensităţii şi direcţiei legăturii.

Cazul cel mai simplu de legătură dintre două variabile X şi Y este cel al regresiei liniare. Aceasta înseamnă că mulţimea punctelor (x_i, y_i) se grupează geometric după modelul unei drepte:

$$y_x = a + b\,x \qquad (14.1)$$

Modelul regresiei liniare corespunde în practică legăturii dintre două variabile care variază în progresie aritmetică.

Introdus în statistică de către biologul şi statisticianul englez Francis Galton[x] (1822-1911), termenul de regresie derivă de la latinescul regressio care înseamnă întoarcere. In cadrul cercetărilor sale de ereditate, Galton a constatat că din părinţi cu talie mai mică decît media pe colectivitate se nasc copii cu o talie superioară mediei colectivităţii şi invers. Deşi în cazul general denumirea este improprie, ea a devenit o noţiune consacrată în statistică.

Geometric, parametrii a şi b din ecuaţia (14.1) au semnificaţia din fig. 14.1.

---

x) Impreună cu K. Pearson este întemeietorul şcolii engleze de statistică matematică şi a publicaţiei ei principale - revista "Biometrika".

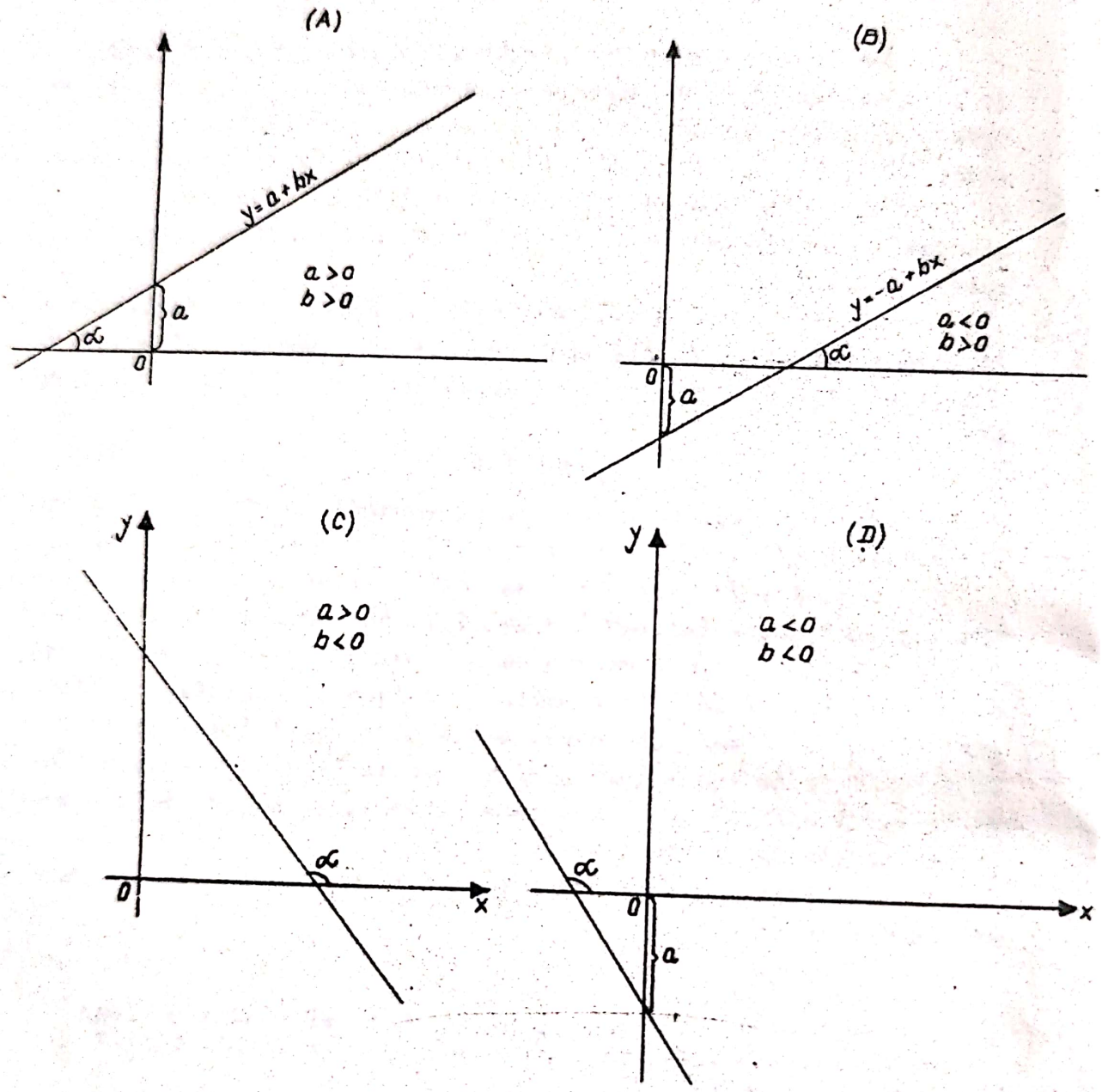


Fig. 14.1. TIPURI DE LEGĂTURI SIMPLE LINIARE

Parametrul a reprezintă <u>ordonata la origine</u> şi poate lua atît valori pozitive cît şi valori negative, iar parametrul b poartă denumirea de coeficient de regresie şi măsoară înclinaţia dreptei faţă de axa absciselor: $b = tg\alpha$

Determinarea parametrilor dreptei de regresie şi în general a oricărei funcţii de regresie se poate face cu ajutorul metodei punctelor selecţionate sau cu ajutorul metodei celor mai mici pătrate.

Se ştie că orice dreaptă este descrisă de ecuaţia (14.1). Determinarea parametrilor acestei funcţii cu ajutorul metodei punctelor selectionate se face pe baza a două puncte situate pe dreapta trasată vizual pe grafic, de regulă cîte unul de fiecare extremitate a segmentului şi apoi se citesc pe grafic coordonatele punctelor respective. Să presupunem că s-au selecţionat punctele de coordonate (x_1,y_1) şi (x_2,y_2). Substituind aceste valori în ecuaţia (14.1) rezultă două ecuaţii:

$$\begin{cases} a + b\,x_1 = y_2 \\ a + b\,x_2 = y_2 \end{cases} \qquad (14.2)$$

Rezolvînd sistemul (14.2) se obţin valorile celor doi parametri a şi b.

Din geometria analitică se cunoaşte ecuaţia unei drepte care trece prin două puncte (x_1, y_1) şi (x_2, y_2) :

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

în care $\frac{y_2 - y_1}{x_2 - x_1}$ reprezintă coeficientul unghiular al dreptei.

Numărul punctelor selecţionate în fiecare caz particular este egal cu numărul parametrilor ce urmează a fi estimaţi şi deci egal cu numărul ecuaţiilor din sistem.

Trasarea vizuală a dreptei depinde de experienţa şi de competenţa cercetătorului şi diferă de la un cercetător la altul, ceea ce înseamnă că şi punctele care se selecţionează de către cercetători diferiţi vor fi diferite, ei vor fi influenţaţi într-o anumită măsură de subiectivism . Datorită acestui fapt metoda punctelor selecţionate are un caracter aproximativ, Subiectivismul cercetătorului ar putea fi diminuat prin împărţirea valorilor variabilelor corelate într-un anumit număr de părţi egale

şi prin înlocuirea valorilor individuale ale coordonatelor punctelor situate pe dreapta trasată visual cu mediile a două grupe de termeni.

Caracterul subiectiv al dreptei de regresie obţinute pe baza metodei punctelor selecţionate se înlătură în totalitate prin utilizarea metodei celor mai mici pătrate.

Determinarea parametrilor dreptei de regresie cu ajutorul metodei celor mai mici pătrate, se bazează pe următorul principiu: suma pătratelor diferenţelor dintre valorile empirice y_i şi valorile teoretice y_{x_i}, date de ecuaţia de regresie, să fie minimă. Geometric, pătratul diferenţelor dintre y_i şi y_{x_i} se exprimă prin aria unor pătrate ale căror laturi sînt egale cu $y_i - y_{x_i}$ (fig.14.2).

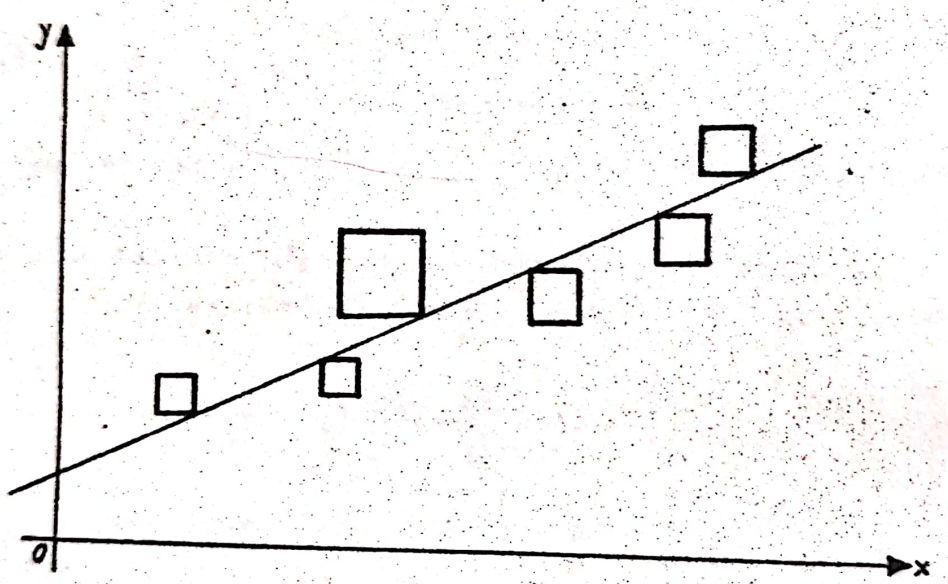


Fig.14.2. CELE MAI MICI PĂTRATE ALE ABATERILOR VALORILOR EMPIRICE FAŢĂ DE VALORILE TEORETICE ALE LUI y

Este clar că suma ariilor acestor pătrate va fi cu atît mai mică cu cît valorile empirice ale lui y se vor apropia mai mult de valorile teoretice corespunzătoare şi invers.

$$S = \sum_{i=1}^{n} (y_i - y_{x_i})^2 = \text{minim} \qquad (14.3)$$

Substituind y_x cu valoarea lui din (14.1), relaţia (14.3) devine:

$$S = \sum_{i=1}^{n} (y_i - a - bx_i)^2 = \text{minim} \qquad (14.4)$$

Minimul funcţiei S se obţine calculînd derivatele de ordinul întîi în raport cu cei doi parametri a şi b şi apoi egalîndu-le cu zero:

$$\frac{\partial S}{\partial a} = 2\sum_{i=1}^{n}(y_i - a - bx_i)(-1) = 0 \qquad (14.5)$$

$$\frac{\partial S}{\partial b} = 2\sum_{i=1}^{n}(y_i - a - bx_i)(-x_i) = 0$$

Prin transformări elementare din relaţiile (14.5) se obţine următorul sistem de ecuaţii, numit sistem de ecuaţii normale:

$$\begin{aligned} n\,a + b\sum x_i &= \sum y_i \\ a\sum x_i + b\sum x_i^2 &= \sum x_i y_i \end{aligned} \qquad (14.6)$$

Rezolvarea sistemului de ecuaţii normale (14.6) ale dreptei de regresie se poate efectua prin diferite metode cum ar fi: metoda substituţiei, metoda lui Cramer, metoda lui Gauss şi altele.

14.1.1. Metode de rezolvare a sistemului de ecuaţii normale liniare

1) **Metoda substituţiei.** Din prima ecuaţie rezultă:

$$a = \frac{\sum y_i - b\sum x_i}{n} = \bar{y} - b\bar{x} \qquad (14.7)$$

Relaţia (14.7) exprimă faptul că dreapta de regresie trece prin punctul de coordonate $(\bar{x},\bar{y})$.

Pe baza relaţiei (14.7), ecuaţia dreptei (14.1) devine:

$$\bar{y}_x = \bar{y} - b\bar{x} + bx_i \qquad (14.8)$$

sau

$$y_x = \bar{y} + b(x_i - \bar{x})$$

Relaţia (14.8) exprimă valorile teoretice ale variabilei dependente y în funcţie de coficientul de regresie b şi de mediile valorilor empirice $\bar{x}$ şi $\bar{y}$ ale celor două variabile.

Substituind pe a din (14.7) în ecuaţia a doua a sistemului (14.6) rezultă:

$$\left(\frac{\sum y_i}{n} - b\frac{\sum x_i}{n}\right)\sum x_i + b\sum x_i^2 = \sum x_i y_i$$

$$\frac{\sum x_i \sum y_i}{n} - \frac{b\left(\sum x_i\right)^2}{n} + b\sum x_i^2 = \sum x_i y_i$$

de unde

$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \qquad (14.9)$$

sau

$$b = \frac{\sum x_i y_i - n\,\bar{x}\,\bar{y}}{\sum x_i^2 - n\bar{x}^2} \qquad (14.10)$$

Dacă se face o mişcare de translaţie a dreptei de regresie în aşa fel încît originea axelor de coordonate să coincidă cu punctul de coordonate $\bar{x}$ şi $\bar{y}$, valoarea coeficientului de regrse b nu se modifică, deoarece înclinaţia dreptelor faţă de axa absciselor rămîne neschimbată.

Astfel, substituind $x_i' = x_i - \bar{x}$ şi $y_i' = y_i - \bar{y}$ în (14.7) rezultă:

$$a = \frac{\sum(y_i - \bar{y}) - b\sum(x_i - \bar{x})}{n} = 0$$

şi

$$b = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \qquad (14.11)$$

Dacă se face o mişcare de translaţie a dreptei de regresie în aşa fel încît originea axelor să se mute în punctul de abscisă $\bar{x}$, atunci sistemul de ecuaţie normală (14.6) devine:

$$\begin{aligned} na + b\sum(x_i - \bar{x}) &= \sum y_i \\ a\sum(x_i - \bar{x}) + b\sum(x_i - \bar{x})^2 &= \sum(x_i - \bar{x})y_i \end{aligned} \qquad (14.12)$$

Deoarece $\sum(x_i - \bar{x}) = 0$ sistemul (14.12) se poate scrie astfel:

$$na = \sum y_i$$
$$b\sum(x_i - \bar{x})^2 = \sum(x_i - \bar{x})y_i$$

Iar formulele de calcul ale parametrilor ecuaţiei regresiei simple liniare vor avea expresiile:

$$a = \bar{y} \qquad (14.13)$$

$$b = \frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2} \qquad (14.14)$$

Să considerăm, pentru exemplificare, legătura dintre producţia de energie termoelectrică şi producţia de cărbune extras în România în perioada 1960 - 1976 (tabelul 14.1).

Tabelul 14.1. Producţia de cărbune extras şi de energie termoelectrică a României în perioada 1960 - 1976

Anul	Cărbune extras (milioane tone)	Energie termoelectrică (miliarde kwh)
1	2	3
1960	8,2	7,3
1961	8,7	8,2
1962	9,6	9,4
1963	10,3	11,1
1964	11,1	13,3
1965	12,1	16,2
1966	13,5	19,8
1967	15,0	23,3
1968	17,0	26,3
1969	19,2	29,3
1970	22,8	32,3
1971	23,0	35,0
1972	25,3	36,1
1973	26,7	39,2
1974	29,2	40,6
1975	29,4	45,0
1976	28,1	50,2

Sursa: Anuarele statistice ale Republicii Socialiste România.

Legătura dintre aceste două variabile este evidentă. O mare parte din producţia de cărbune este utilizată în termocentrale pentru producţia de energie termoelectrică, dar cărbunele nu este singurul combustibil utilizat în termocentrale.

Legătura dintre producţie şi energie termoelectrică (variabilă rezultativă) şi cantitatea de cărbune extras (variabilă independentă) se prezintă sugestiv în fig. 14.3.

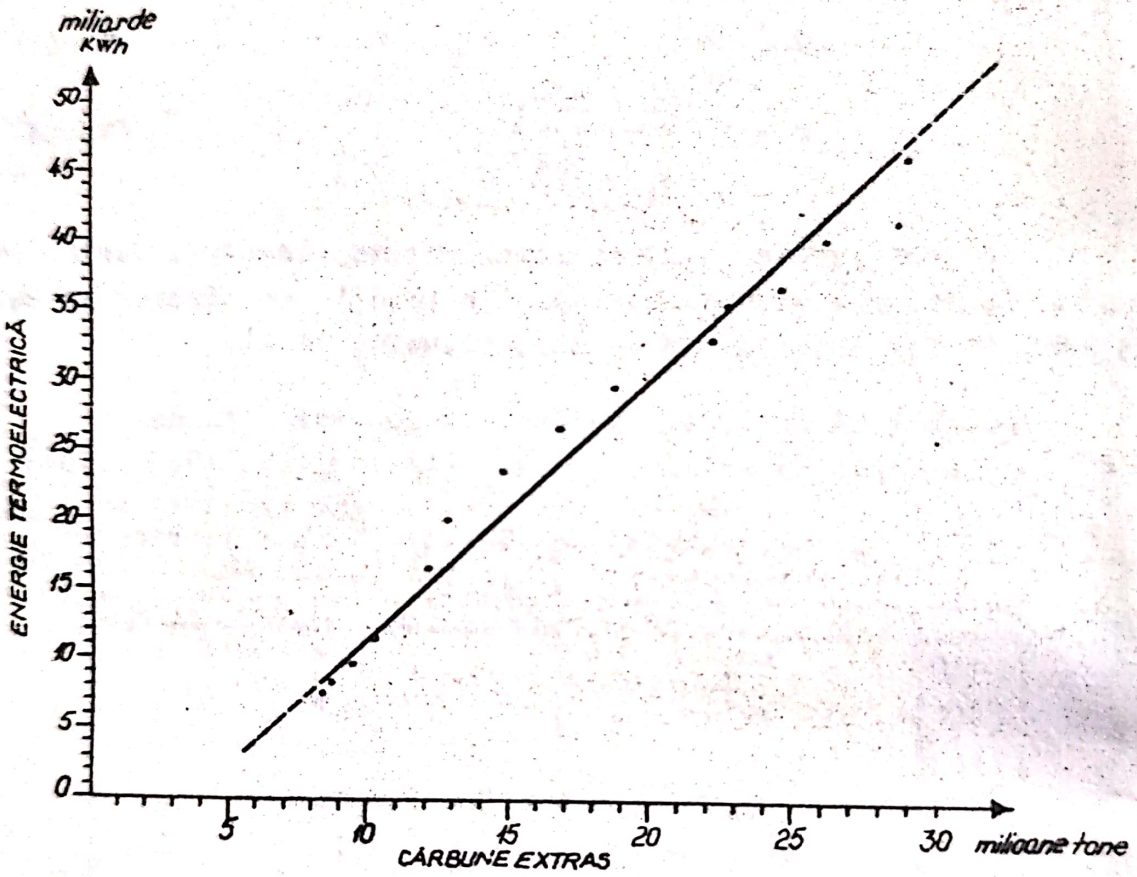


Fig. 14.3. CORELAŢIA DINTRE PRODUCŢIE DE CĂRBUNE ŞI PRODUCŢIE DE ENERGIE TERMOELECTRICĂ

Pentru aplicarea formulelor (14.13) şi (14.14), în vederea determinării parametrilor ecuaţiei liniare de regresie, calculele se sistematizează cu ajutorul tabelului 14.2.

Substituind datele din tabelul 14.2 în formulele (14.13) şi (14.14) rezultă:

$$a = \bar{y} = \frac{\sum y_i}{n} = \frac{442,6}{17} = 26,35 \text{ miliarde kwh}$$

iar

$$b = \frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2} = \frac{1677,205}{961,318} = 1,744$$

Inlocuind succesiv valorile parametrilor a şi b, de mai sus, în ecuaţia de regresie (14.8) se obţin valorile teoretice ale variabilei dependente y_x (coloana 6 a tabelului 14.2).

Valorile teoretice ale variabilei dependente s-au reprezentat în fig. 14.3, de unde se observă că dreapta de regresie se apropie în mare măsură de valorile empirice, ceea ce înseamnă că modelul liniar de regresie descrie în mod corespunzător legătura dintre cele două variabile.

Tabelul 14.2. Determinarea parametrilor ecuaţiei liniare de regresie prin metoda substituţiei

Anul	Cărbune extras (mil.t.) (x_i)	Energie termoelectrică (miliarde kwh(y_i)	$x_i-\bar{x}$	$(x_i-\bar{x})y_i$	$(x_i-\bar{x})^2$	y_x
0	1	2	3	4	5	6
1960	8,2	7,3	-9,988	-72,912	99,760	8,616
1961	8,7	8,2	-9,488	-77,802	90,022	9,488
1962	9,6	9,4	-8,588	-80,727	73,754	11,058
1963	10,3	11,1	-7,888	-87,557	62,221	12,278
1964	11,1	13,3	-7,088	-94,270	50,240	13,674
1965	12,1	16,2	-6,088	-98,626	37,064	15,418
1966	13,5	19,8	-4,688	-92,822	21,977	17,859
1967	15,0	23,3	-3,188	-74,280	10,163	20,475
1968	17,0	26,3	-1,188	-31,244	1,411	23,963
1969	19,2	29,3	-1,012	29,652	1,024	27,800
1970	22,8	32,3	4,612	148,968	21,271	34,078
1971	23,0	35,0	4,812	168,420	23,155	34,427
1972	25,3	36,1	7,112	256,743	50,581	38,438
1973	26,7	39,2	8,512	333,670	72,454	40,880
1974	29,2	40,6	11,012	447,087	121,264	45,240
1975	29,4	45,0	11,212	504,540	125,709	45,589
1976	28,1	50,2	9,912	497,582	98,248	43,322
TOTAL	309,2	442,6	0	1677,205	961,318	442,603

Verificarea exactităţii calculelor ecuaţiei liniare de regresie se poate efectua pe baza relaţiei care există între valorile reale ale variabilei dependente y_i şi valorile teoretice y_{x_i} ale acestei variabile şi anume suma valorilor variabilei empirice este egală cu suma valorilor teoretice, adică:

$$\sum y_i = \sum y_{x_i}$$

In exemplul nostru, totalul pe coloana 2 a tabelului 14.2 este aproximativ egal cu totalul de pe coloana 6 a aceluiaşi tabel. Diferenţa de 0,3‰ dintre aceste sume se datorează numărului limitat de zecimale luat în calcule.

2) <u>Metoda lui Cramer</u>. Aplicarea acestei metode presupune calculul şi analiza prealabilă a determinantului principal şi a determinanţilor caracteristici corespunzători celor doi parametri.

Valoarea determinantului principal, obţinut pe baza sistemului (14.6), se determină astfel:

$$\Delta = \begin{vmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{vmatrix} = \left[n\sum x_i^2 - (\sum x_i)^2 \right] > 0$$

Valorile determinanţilor caracteristici vor fi:

$$\Delta_a = \begin{vmatrix} \sum y_i & \sum x_i \\ \sum x_i y_i & \sum x_i^2 \end{vmatrix} = \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i$$

$$\Delta_b = \begin{vmatrix} n & \sum y_i \\ \sum x_i & \sum x_i y_i \end{vmatrix} = n\sum x_i y_i - \sum x_i \sum y_i$$

Aplicînd regula lui Cramer, pentru rezolvarea sistemului de ecuaţii normale (14.6), rezultă următoarele formule pentru parametri ecuaţiei de regresie :

$$a = \frac{\Delta_a}{\Delta} = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2} \qquad (14.15)$$

$$b = \frac{\Delta_b}{\Delta} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} \qquad (14.16)$$

Scanned with OKEN Scanner

Pentru exemplificare, considerăm datele din tabelul 14.1. Calculele intermediare pentru determinarea parametrilor ecuației regresiei simple liniare pe baza formulelor (14.15) și 14.16) se prezintă sistematizat în tabelul 14.3.

Înlocuind în formulele (14.15) și (14.16) elementele de calcul din tabelul 14.3, rezultă:

$$a = \frac{6584,12.442,6 - 309,2.9726,46}{17.658,12 - 309,2^2} = \frac{-93289,9}{16325,4} = -5,7144$$

și

$$b = \frac{17.9726,46 - 309,2.442,6}{17.6584,12 - 309,2^2} = \frac{28497,9}{16325,4} = 1,7456$$

Substituind valorile parametrilor a și b în formula ecuației de regresie rezultă valorile teoretice ale variabilei dependente y (coloana 6 din tabelul 14.3):

$$y_{x_i} = -5,7144 + 1,7456\, x_i$$

Tabelul 14.3. Calculul parametrilor ecuației liniare de regresie pe baza metodei lui Cramer

Anul	x_i	y_i	x_i^2	y_i^2	$x_i y_i$	y_x	x_y
0	1	2	3	4	5	6	7
1960	8,2	7,3	67,24	53,29	59,86	8,599	7,837
1961	8,7	8,2	75,69	67,24	71,34	9,972	8,333
1962	9,6	9,4	92,16	88,36	90,24	11,043	8,996
1963	10,3	11,1	106,09	123,21	114,33	12,265	9,934
1964	11,1	13,3	123,21	176,89	147,63	13,662	11,149
1965	12,1	16,2	146,41	262,44	196,02	15,407	12,749
1966	13,5	19,8	182,25	392,04	267,30	17,851	14,737
1967	15,0	23,3	225,00	542,89	349,50	20,470	16,669
1968	17,0	26,3	289,00	691,69	447,10	23,961	18,325
1969	19,2	29,3	368,64	858,49	562,56	27,801	19,981
1970	22,8	32,3	519,84	1043,29	736,44	34,085	21,637
1971	23,0	35,0	529,00	1225,00	805,00	34,434	23,127
1972	25,3	36,1	640,09	1303,21	913,33	38,449	23,734
1973	26,7	39,2	712,89	1536,64	1046,64	40,893	25,445
1974	29,2	40,6	852,64	1648,36	1185,52	45,257	26,218

Tabelul 14.3 (continuare)

0	1	2	3	4	5	6	7
1975	29,4	45,0	864,36	2025,00	1323,00	45,606	28,64
1976	28,1	50,2	789,61	2520,04	1410,62	43,337	31,51
TOTAL	309,2	442,6	6584,12	14558,08	9726,46	442,592	309,03

Ca şi în cazul metodei substituţiei se verifică şi acum relaţia: $\sum y_i = \sum y_{x_i}$

Coeficientul de regresie b = 1,7456 are o valoare apropiată de cea obţinută pe baza rezolvării sistemului de ecuaţii normale cu ajutorul metodei substituţiei şi arată că între cantitatea de cărbune extras şi producţia de energie termoelectrică există o legătură directă şi anume la o creştere a producţiei de cărbune cu un milion tone producţia de energie electrică creşte cu 1,74 miliarde kwh.

3. <u>Metoda lui Gauss</u> se bazează pe o succesiune de transformări elementare asupra unei matrici pătratice în aşa fel încît matricea iniţială se transformă într-o matrice unitate. Aceasta înseamnă că transformarea matricii A_{nxn} într-o matrice I_{nxn} se poate obţine cu ajutorul unor succesiuni de transformări liniare care dau acelaşi rezultat ca şi înmulţirea matricii pătratice cu inversa ei:

$$A^{-1}_{nxn} \cdot A_{nxn} = I_{nxn} \qquad (14.17)$$

Invers, dacă aceleaşi transformări se aplică unei matrici unitate se va obţine matricea inversă A^{-1}_{nn}.

Pe baza acestui principiu se pot rezolva sisteme de ecuaţii în care numărul ecuaţiilor este egal cu numărul necunoscutelor:

$$A_{nxn} \cdot X_{nxl} = P_{nxl} \qquad (14.18)$$

Soluţia X^H se obţine formînd matricea extinsă (A P) şi efectuînd transformări liniare care vor reduce partea din stînga A_{nxn} a acestei matrici la o matrice unitate I_{nxn}. In acelaşi timp, aceste transformări se vor aplica şi asupra părţii din dreapta P_{nxl}, deoarece transformările liniare vor fi aplicate asupra tuturor celor n+1 elemente de pe fiecare linie. Rezultatul acestor transformări (A P) $\longrightarrow$ (I X^H) este acelaşi ca şi cum ar fi înmulţit P cu matricea inversă A^{-1}, adică $A^{-1}P = X^H$.

Tabelul 14.3 (continuare)

0	1	2	3	4	5	6	7
1975	29,4	45,0	864,36	2025,00	1323,00	45,606	28,64
1976	28,1	50,2	789,61	2520,04	1410,62	43,337	31,51
TOTAL	309,2	442,6	6584,12	14558,08	9726,46	442,592	309,03

Ca şi în cazul metodei substituţiei se verifică şi acum relaţia: $\sum y_i = \sum y_{x_i}$

Coeficientul de regresie b = 1,7456 are o valoare apropiată de cea obţinută pe baza rezolvării sistemului de ecuaţii normale cu ajutorul metodei substituţiei şi arată că între cantitatea de cărbune extras şi producţia de energie termoelectrică există o legătură directă şi anume la o creştere a producţiei de cărbune cu un milion tone producţia de energie electrică creşte cu 1,74 miliarde kwh.

3. Metoda lui Gauss se bazează pe o succesiune de transformări elementare asupra unei matrici pătratice în aşa fel încît matricea iniţială se transformă într-o matrice unitate. Aceasta înseamnă că transformarea matricii A_{nxn} într-o matrice I_{nxn} se poate obţine cu ajutorul unor succesiuni de transformări liniare care dau acelaşi rezultat ca şi înmulţirea matricii pătratice cu inversa ei:

$$A_{nxn}^{-1} \cdot A_{nxn} = I_{nxn} \qquad (14.17)$$

Invers, dacă aceleaşi transformări se aplică unei matrici unitate se va obţine matricea inversă A_{nn}^{-1}.

Pe baza acestui principiu se pot rezolva sisteme de ecuaţii în care numărul ecuaţiilor este egal cu numărul necunoscutelor:

$$A_{nxn} \cdot X_{nxl} = P_{nxl} \qquad (14.18)$$

Soluţia X^H se obţine formînd matricea extinsă (A P) şi efectuînd transformări liniare care vor reduce partea din stînga A_{nxn} a acestei matrici la o matrice unitate I_{nxn}. In acelaşi timp, aceste transformări se vor aplica şi asupra părţii din dreapta P_{nxl}, deoarece transformările liniare vor fi aplicate asupra tuturor celor n+1 elemente de pe fiecare linie. Rezultatul acestor transformări (A P) $\longrightarrow$ (I X^H) este acelaşi ca şi cum ar fi înmulţit P cu matricea inversă A^{-1}, adică $A^{-1}P = X^H$.

Tabelul 14.3 (continuare)

p	1	2	3	4	5	6	7
1975	29,4	45,0	864,36	2025,00	1323,00	45,606	28,647
1976	28,1	50,2	789,61	2520,04	1410,62	43,337	31,517
TOTAL	309,2	442,6	6584,12	14558,08	9726,46	442,592	309,035

Ca şi în cazul metodei substituţiei se verifică şi acum relaţia: $\sum y_i = \sum y_{x_i}$

Coeficientul de regresie b = 1,7456 are o valoare apropiată de cea obţinută pe baza rezolvării sistemului de ecuaţii normale cu ajutorul metodei substituţiei şi arată că între cantitatea de cărbune extras şi producţia de energie termoelectrică există o legătură directă şi anume la o creştere a producţiei de cărbune cu un milion tone producţia de energie electrică creşte cu 1,74 miliarde kwh.

3. <u>Metoda lui Gauss</u> se bazează pe o succesiune de transformări elementare asupra unei matrici pătratice în aşa fel încît matricea iniţială se transformă într-o matrice unitate. Aceasta înseamnă că transformarea matricii A_{nxn} într-o matrice I_{nxn} se poate obţine cu ajutorul unor succesiuni de transformări liniare care dau acelaşi rezultat ca şi înmulţirea matricii pătratice cu inversa ei:

$$A^{-1}_{nxn} \cdot A_{nxn} = I_{nxn} \qquad (14.17)$$

Invers, dacă aceleaşi transformări se aplică unei matrici unitate se va obţine matricea inversă A^{-1}_{nn}.

Pe baza acestui principiu se pot rezolva sisteme de ecuaţii în care numărul ecuaţiilor este egal cu numărul necunoscutelor:

$$A_{nxn} \cdot X_{nxl} = P_{nxl} \qquad (14.18)$$

Soluţia X^H se obţine formînd matricea extinsă (A P) şi efectuînd transformări liniare care vor reduce partea din stînga A_{nxn} a acestei matrici la o matrice unitate I_{nxn}. In acelaşi timp, aceste transformări se vor aplica şi asupra părţii din dreapta P_{nxl}, deoarece transformările liniare vor fi aplicate asupra tuturor celor n+1 elemente de pe fiecare linie. Rezultatul acestor transformări (A P) $\longrightarrow$ (I X^H) este acelaşi ca şi cum ar fi înmulţit P cu matricea inversă A^{-1}, adică $A^{-1}P = X^H$.

Să aplicăm această metodă pentru rezolvarea sistemului de ecuaţii normale (14.6):

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix} \qquad (14.19)$$

Matricea extinsă va fi:

$$\begin{pmatrix} n & \sum x_i & \sum y_i \\ \sum x_i & \sum x_i^2 & \sum x_i y_i \end{pmatrix}$$

Se urmăreşte reducerea primelor două coloane ale acestei matrici la o matrice unitate (2x2) prin transformări pe linii. O metodă sistematică necesită transformarea la început a elementului (1,1) în 1, iar celelalte elemente de pe aceeaşi coloană în zero. Apoi se transformă următorul element de pe diagonala principală (2,2) în unitate şi celelalte elemente de pe coloana 2 în zero. In continuare transformările se vor aplica asupra elementului (3,3) şi aşa mai departe, pînă cînd elementul (n,n) devine egal cu 1 şi celelalte elemente de pe coloana n devin zero.

In cazul nostru, n = 2 , deci procedeul se va repeta numai de două ori. Schematic, succesiunea transformărilor şi a rezultatelor în fiecare treaptă se prezintă astfel:

Matricea extinsă $\begin{pmatrix} n & \sum x_i & \sum y_i \\ \sum x_i & \sum x_i^2 & \sum x_i y_i \end{pmatrix}$.

Treapta I : se înmulţeşte linia 1 cu $\frac{1}{n}$:

$$\begin{pmatrix} 1 & \frac{\sum x_i}{n} & \frac{\sum y_i}{n} \\ \sum x_i & \sum x_i^2 & \sum x_i y_i \end{pmatrix}$$

Treapta 2 : se adună la linia 2 scalarul $(-\sum x_i)$ înmulţit cu linia 1 :

$$\begin{pmatrix} 1 & \frac{\sum x_i}{n} & \frac{\sum y_i}{n} \\ 0 & \sum x_i^2 - \frac{(\sum x_i)^2}{n} & \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} \end{pmatrix}$$

Treapta 3: se înmulţeşte linia 2 cu scalarul $\frac{1}{\sum(x_i - \bar{x})^2}$ şi se simplifică expresia folosind notaţiile pentru medii şi momente simple:

$$\begin{pmatrix} 1 & \bar{x} & \bar{y} \\ 0 & 1 & \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \end{pmatrix}.$$

Treapta 4: se adună la linia 1 scalarul $(-\bar{x})$ înmulţit cu linia a doua:

$$\begin{pmatrix} 1 & 0 & \bar{y} - \frac{\bar{x}\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \\ 0 & 1 & \frac{\sum(x_i - \bar{x})(y_i - y)}{\sum(x_i - \bar{x})^2} \end{pmatrix}$$

Deci sistemul de ecuaţii normale iniţial s-a transformat într-un sistem echivalent :

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \quad \begin{pmatrix} \bar{y} - b\bar{x} \\ \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \end{pmatrix}$$

De unde rezultă:

$$a = \bar{y} - b\bar{x}$$

şi

$$b = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

Scanned with OKEN Scanner

Menţionăm că aceste formule sînt identice cu cele obţinute anterior prin aplicarea metodei substituţiei (vezi 14.7 şi 14.11).

Pentru ilustrarea metodei lui Gauss de rezolvare a unui sistem de ecuaţii liniare să substituim în (14.6) elementele de calcul din tabelul 14.3., astfel:

$$17\,a + 309{,}2\,b = 442{,}6$$
$$309{,}2\,a + 6584{,}12\,b = 9726{,}46$$

Tabelul 14.4. Metoda lui Gauss de rezolvare a sistemelor de ecuaţii liniare

	a	b	
a	17	309,2	442,6
b	309,2	6584,12	9726,46
a	1	18,188235	26,035294
b	0	961,318	1676,3471
a	1	0	-5,714
b	0	1	1,744

Din ultima iteraţie rezultă:

$$a = -5{,}714$$
$$b = 1{,}744$$

Deci ecuaţia de regresie care descrise legătura dintre producţia de energie electrică şi cantitatea de cărbune extras va fi de forma:

$$Y_x = -5{,}714 + 1{,}744\,x_i$$

14.1.2. Dualitatea în regresie

In cazul a două variabile există două drepte de regresie y_x şi x_y, care (determinate cu ajutorul metodei celor mai mici pătrate) au coeficienţi unghiulari diferiţi, astfel:

- pentru dreapta de regresie $y_x = a_1 + b_1 x_i$ coeficientul de regresie are expresia:

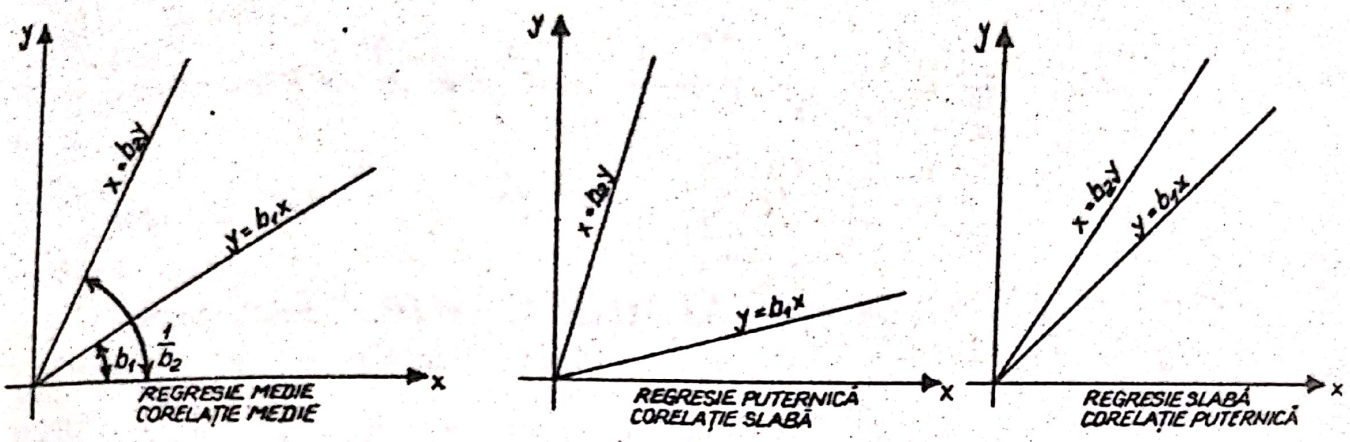


Fig.14.5. CORESPONDENȚA DINTRE REGRESIE ȘI CORELAȚIE

Dacă dreapta de regresie trece prin origine (fig. 14.4 și fig. 14.5) atunci ea depinde numai de un singur parametru și anume de coeficientul de regresie b:

$$y_x = b_1 x \qquad \text{și} \qquad x_y = b_2 y$$

Și în acest caz particular, coeficientul de regresie se determină tot cu ajutorul metodei celor mai mici pătrate. De exemplu, pentru determinarea lui b_1 se va proceda astfel:

$$S = \sum (y_i - y_x)^2 = \text{minim}$$

Substituind pe y_x cu $b_1 x$ rezultă:

$$S = \sum (y_i - b_1 x_i)^2 = \text{minim}$$

Calculînd derivata a întîia a lui S în raport cu b și egalînd-o cu zero se obține:

$$\frac{\partial S}{\partial b} = 2 \sum (y_i - b_1 x_i)(-x_i) = 0$$

și

$$\sum y_i x_i = b_1 \sum x_i^2$$

de unde

$$b_1 = \frac{\sum x_i y_i}{\sum x_i^2} \qquad (14.22)$$

In mod analog se determină coeficientul de regresie b_2 şi se obţine:

$$b_2 = \frac{\sum x_i y_i}{\sum y_i^2} \qquad (14.23)$$

Coeficientul b_2 se calculează în practică numai în cazul în care se poate considera că între variabilele x şi y există o interacţiune reciprocă reală.

In cazul general al regresiei liniare cînd a $\neq$ 0, parametrii dreptei de regresie $x_y = a + by$ se determină tot cu ajutorul metodei celor mai mici pătrate, astfel:

$$S = \sum (x_i - x_y)^2 = \text{minim}$$

$$S = \sum (x_i - a - by)^2 = \text{minim}$$

Efectuînd derivatele parţiale ale funcţiei S în raport cu cei doi parametri şi egalîndu-le cu zero, se obţine sistemul de ecuaţii normale:

$$na + b\sum y_i = \sum x_i$$

$$a\sum y_i + b\sum y_i^2 = \sum y_i x_i$$

de unde, prin rezolvare, rezultă:

$$a = \frac{\sum x_i \sum y_i^2 - \sum y_i \sum y_i x_i}{n\sum y_i^2 - (\sum y_i)^2}$$

$$b = \frac{n\sum y_i x_i - \sum y_i \sum x_i}{n\sum y_i^2 - (\sum y_i)^2} \qquad (14.24)$$

Să considerăm, pentru exemplificarea dualităţii în regresie, legătura dintre cantitatea de cărbune extras şi producţia de energie termoelectrică (tabelul 14.1).

Scanned with OKEN Scanner

Parametrii ecuației de regresie a variabilei x în funcție de variabila Y se determină cu ajutorul formulelor (14.23) și (14.24):

$$a_2 = \frac{309,2 \cdot 14558,08 - 442,6 \cdot 9726,46}{17 \cdot 14558,08 - 442,6^2} = 3,807$$

$$b_2 = \frac{17 \cdot 9726,46 - 309,2 \cdot 442,6}{17 \cdot 14558,08 - 442,6^2} = 0,552$$

$a_2 = 3,807$ exprimă faptul că dreapta de regresie x_y intersectează axa x-lor în punctul de coordonate $y = 0$ și $x = 3,807$ miliarde tone.

$b_2 = 0,552$ miliarde tone, înseamnă că producția de cărbune crește cu peste jumătate de milion de tone atunci cînd producția de energie termoelectrică crește cu un miliard de kwh. Ecuația de regresie va fi:

$$x_y = 3,807 + 0,552y_i$$

Valorile teoretice ale lui x_y corespunzătoare diferitelor valori ale variabilei Y s-au înscris în coloana 7 a tabelului 14.3.

Dacă perechile de valori x_i, y_i se întîlnesc în cadrul distribuției de mai multe ori "n_i", atunci cele două variabile vor apare ponderate în cadrul sistemului de ecuații normale:

$$\begin{aligned} a\sum n_i + b\sum x_i n_i &= \sum y_i n_i \\ a\sum x_i n_i + b\sum x_i^2 n_i &= \sum x_i y_i n_i \end{aligned} \qquad (14.25)$$

Rezolvînd sistemul (14.25) se obțin formulele de calcul ale celor doi parametri:

$$a = \frac{\sum x_i^2 n_i \sum y_i n_i - \sum x_i n_i \sum x_i y_i n_i}{\sum n_i \sum x_i^2 n_i - (\sum x_i n_i)^2} \qquad (14.26)$$

$$b = \frac{\sum n_i \sum x_i y_i n_i - \sum x_i n_i \sum y_i n_i}{\sum n_i \sum x_i^2 n_i - (\sum x_i n_i)^2} \qquad (14.27)$$

Pentru determinarea practică a parametrilor ecuației de regresie y_x, calculele se pot sintetiza sub forma tabelului 14.5.

Tabelul 14.5. Calcule intermediare ale parametrilor ecuației de regresie

Nr. crt.	x_i	y_i	n_i	$x_i n_i$	$y_i n_i$	$x_i^2 n_i$	$x_i y_i n_i$	$y_i^2 n_i$
1	x_1	y_1	n_1	$x_1 n_1$	$y_1 n_1$	$x_1^2 n_1$	$x_1 y_1 n_1$	$y_1^2 n_1$
2	x_2	y_2	n_2	$x_2 n_2$	$y_2 n_2$	$x_2^2 n_2$	$x_2 y_2 n_2$	$y_2^2 n_2$
3	x_3	y_3	n_3	$x_3 n_3$	$y_3 n_3$	$x_3^2 n_3$	$x_3 y_3 n_3$	$y_3^2 n_3$
...	...	...	...	...	...	...	...	...
i	x_i	y_i	n_i	$x_i n_i$	$y_i n_i$	$x_i^2 n_i$	$x_i y_i n_i$	$y_i^2 n_i$
...	...	...	...	...	...	...	...	...
k	x_k	y_k	n_k	$x_k n_k$	$y_k n_k$	$x_k^2 n_k$	$x_k y_k n_k$	$y_k^2 n_k$
Total	-	-	$\sum n_i$	$\sum x_i n_i$	$\sum y_i n_i$	$\sum x_i^2 n_i$	$\sum x_i y_i n_i$	$\sum y_i^2 n_i$

In cazul regresiei variabileiX față de Y sistemul de ecuații normale (14.25) se scrie astfel:

$$\begin{aligned} a\sum n_i + b\sum y_i n_i &= \sum x_i n_i \\ a\sum y_i n_i + b\sum y_i^2 n_i &= \sum y_i x_i n_i \end{aligned} \qquad (14.28)$$

Din sistemul (14.28) prin rezolvare rezultă formulele de calcul ale celor doi parametri:

$$a = \frac{\sum x_i n_i \sum y_i^2 n_i - \sum y_i n_i \sum y_i x_i n_i}{\sum n_i \sum y_i^2 n_i - (\sum y_i n_i)^2} \qquad (14.29)$$

$$b = \frac{\sum n_i \sum y_i x_i n_i - \sum y_i n_i \sum x_i n_i}{\sum n_i \sum y_i^2 n_i - (\sum y_i n_i)^2} \qquad (14.30)$$

Sistematizarea calculelor parametrilor ecuației de regresie $x_y = a + by$ se prezintă tot în tabelul 14.5.

Aşa cum am văzut în prima parte a acestei cărţi, în cazul unui volum mare de valori ale celor două variabile, pentru a reduce volumul calculelor se recurge la gruparea variantelor caracteristicilor pe intervale de variaţie. Gruparea, pe lîngă avantajele ei incontestabile, prezintă după se ştie şi dezavantajul că determină o anumită pierdere de informaţie. De aceea, valorile indicatorilor de regresie şi corelaţie calculaţi pe baza datelor negrupate diferă de cei care se obţin cu ajutorul datelor grupate. Diferenţa dintre rezultate se micşorează pe măsură ce distribuţiile empirice, obţinute pe baza grupării, se apropie tot mai mult de o distribuţie normală. Subliniem faptul că este, după unii autori, discutabilă calcularea indicatorilor de regresie şi corelaţie pe baza unor date grupate pe intervale neegale de variaţie.

Dacă numărul perechilor de valori x_i, y_i este mare, atunci pentru simplificarea calculelor este de preferat sistematizarea datelor iniţiale sub forma unui tabel cu dublă intrare. In acest caz sistemul de ecuaţii normale (14.6) devine:

$$\begin{aligned} a\sum\sum n_{ij} + b\sum x_i n_{i\cdot} &= \sum y_j n_{\cdot j} \\ a\sum x_i n_{i\cdot} + b\sum x_i^2 n_{i\cdot} &= \sum\sum x_i y_j n_{ij} \end{aligned} \qquad (14.31)$$

Rezolvînd sistemul (14.31) se obţin formulele uzuale de calcul ale parametrilor ecuaţiei de regresie $y_x = a + bx$:

$$a = \frac{\sum x_i^2 n_{i\cdot} \sum y_j n_{\cdot j} - \sum x_i n_{i\cdot} \sum\sum x_i y_j n_{ij}}{\sum\sum n_{ij} \sum x_i^2 n_{i\cdot} - (\sum x_i n_{i\cdot})^2} \qquad (14.32)$$

$$b = \frac{\sum\sum n_{ij} \sum\sum x_i y_j n_{ij} - \sum x_i n_{i\cdot} \sum y_j n_{\cdot j}}{\sum\sum n_{ij} \sum x_i^2 n_{i\cdot} - (\sum x_i n_{i\cdot})^2} \qquad (14.33)$$

In cazul regresiei $x_y = a + by$, pe baza unui tabel de corelaţie, sistemul de ecuaţii normale va fi:

$$\begin{aligned} a\sum\sum n_{ij} + b\sum y_j n_{\cdot j} &= \sum x_i n_{i\cdot} \\ a\sum y_j n_{\cdot j} + b\sum y_j^2 n_{\cdot j} &= \sum\sum y_j x_i n_{ij} \end{aligned} \qquad (14.34)$$

Tabelul 14.6. Calculele intermediare ale parametrilor ecuației de regresie în cazul unui tabel de corelație

x \ y	y_1	y_2	...	y_j	...	y_s	Total $(n_{i\cdot})$	$x_i n_{i\cdot}$	$x_i^2 n_{i\cdot}$	$\sum_i y_j x_i n_{ij}$	y_{x_i}
x_1	n_{11}	n_{12}	...	n_{ij}	...	n_{1s}	$n_{1\cdot}$	$x_1 n_{1\cdot}$	$x_1^2 n_{1\cdot}$	$\sum y_j x_1 n_{1j}$	y_{x_1}
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2s}	$n_{2\cdot}$	$x_2 n_{2\cdot}$	$x_2^2 n_{2\cdot}$	$\sum y_j x_2 n_{2j}$	y_{x_2}
...	...	...	...	...	...	...	...	...	...	...	...
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{is}	$n_{i\cdot}$	$x_i n_{i\cdot}$	$x_i^2 n_{i\cdot}$	$\sum y_j x_i n_{ij}$	y_{x_i}
...	...	...	...	...	...	...	...	...	...	...	...
x_k	n_{k1}	n_{k2}	...	n_{kj}	...	n_{ks}	$n_{k\cdot}$	$x_k n_{k\cdot}$	$x_k^2 n_{k\cdot}$	$\sum y_j x_k n_{kj}$	y_{x_k}
Total $(n_{\cdot j})$	$n_{\cdot 1}$	$n_{\cdot 2}$	...	$n_{\cdot j}$	...	$n_{\cdot s}$	$n_{\cdot\cdot}$	$\sum x_i n_{i\cdot}$	$\sum x_i^2 n_{i\cdot}$	$\sum\sum y_j x_i n_{ij}$	-
$y_j n_{\cdot j}$	$y_1 n_{\cdot 1}$	$y_2 n_{\cdot 2}$	...	$y_j n_{\cdot j}$	...	$y_s n_{\cdot s}$	$\sum y_j n_{\cdot j}$				
$y_j^2 n_{\cdot j}$	$y_1^2 n_{\cdot j}$	$y_2^2 n_{\cdot 2}$	...	$y_j^2 n_{\cdot s}$	...	$y_s^2 n_{\cdot s}$	$\sum y_j^2 n_{\cdot j}$				
$\sum x_i y_j n_{ij}$	$\sum x_i y_1 n_{i1}$	$\sum x_i y_2 n_{12}$	...	$\sum x_i y_j n_{ij}$	...	$\sum x_i y_s n_{is}$	$\sum\sum x_i y_j n_{ij}$				
x_{y_j}	x_{y_1}	x_{y_2}	...	x_{y_j}	...	x_{y_s}	-				

Scanned with OKEN Scanner

Formulele celor doi parametri a şi b au următoarele expresii:

$$a = \frac{\sum x_i n_{i.} \sum y_j^2 n_{.j} - \sum y_j n_{.j} \sum \sum y_j x_i n_{ij}}{\sum \sum n_{ij} \sum y_j^2 n_{.j} - (\sum y_j n_{.j})^2} \qquad (14.35)$$

$$b = \frac{\sum \sum n_{ij} \sum \sum y_j x_i n_{ij} - \sum x_i n_{i.} \sum y_j n_{.j}}{\sum \sum n_{ij} \sum y_j^2 n_{.j} - (\sum y_j n_{.j})^2} \qquad (14.36)$$

Sistematizat, calculele preliminare necesare determinării parametrilor ecuaţiei de regresie se prezintă în tabelul 14.6.

Să considerăm, pentru exemplificarea calculului parametrilor ecuaţiei de regresie liniară, în cazul sistematizării celor două variabile sub forma unui tabel cu dublă intrare, distribuţia judeţelor României după valoarea fondurilor fixe şi după numărul mediu al personalului. (Tabelul 14.7).

Tabelul 14.7. Distribuţia judeţelor după valoarea fondurilor fixe şi numărul mediu al personalului în anul 1976

Mii persoane / Milioane lei fond.fixe	-50	50-100	100-150	150-200	200 -	TOTAL
- 10	3	3	-	-	-	6
10 - 20	-	8	3	-	-	11
20 - 30	-	4	6	2	-	12
30 - 40	-	-	2	1	1	4
40 -	-	-	-	3	3	6
TOTAL	3	15	11	6	4	39

Sursa: Anuarul statistic al Republicii Socialiste România, 1977, D.C.S., p. 93 şi 104 - 105

Tabelul 14.8. Calculul parametrilor regresiei simple liniare în cazul unei distribuţii bidimensionale

	X \ Y	-50	50-100	100-150	150-200	200-	TOTAL ($n_{i.}$)	x_i	$x_i n_{i.}$	$x_i^2 n_{i.}$	$\sum_j x_i y_j n_{ij}$	y_{x_i}
0	1	2	3	4	5	6	7	8	9	10	11	12
1	- 10	3	3	-	-	-	6	5	30	150	1500	48,59
2	10 - 20	-	8	3	-	-	11	15	165	2475	14625	85,63
3	20 - 30	-	4	6	2	-	12	25	300	7500	85000	122,67
4	30 - 40	-	-	2	1	1	4	35	140	4900	22750	159,72
5	40 -	-	-	-	3	3	6	45	270	12150	54000	196,76
6	Total ($n_{.j}$)	3	15	11	6	4	39	-	905	27175	127875	-
7	y_j	25	75	125	175	225	-					
8	$y_j n_{.j}$	75	1125	1375	1050	900	4525					
9	$y_j^2 n_{.j}$	1875	28125	103125	131250	202500	466875					
10	$\sum_i x_i y_j n_{ij}$	375	17625	33125	38500	38250	127875					

Scanned with OKEN Scanner

Calculele intermediare necesare determinării parametrilor ecuaţiilor de regresie se prezintă în mod sistematizat în tabelul 14.8. Rîndul y_j şi respectiv coloana x_i reprezintă mijloacelor intervalelor de variaţie ale celor două variabile. Rîndul 10 al tabelului 14.8 s-a determinat înmulţind fiecare valoare a lui y_j cu toate valorile lui x_i şi cu frecvenţele de la intersecţia lui y_j cu rîndul lui x_i adică cu n_{ij} , astfel:

$$\sum_{i=1}^{5} y_1 x_i n_{i1} = 25.5.3+25.15.0+25.25.0+25.35.0+25.45.0 = 375$$

$$\sum_{i=1}^{5} y_2 x_i n_{i2} = 75.5.3+75.15.8+75.25.4 = 17625$$

$$\sum_{i=1}^{5} y_3 x_i n_{i3} = 125.15.3+125.25.6+125.35.2 = 33125$$

$$\sum_{i=1}^{5} y_4 x_i n_{i4} = 175.25.2+175.35.1+175.45.3 = 38500$$

$$\sum_{i=1}^{5} y_5 x_i n_{i5} = 225.35.1+225,45.3 = 38250$$

$$\sum_{j=1}^{5}\sum_{i=1}^{5} x_i y_j n_{ij} = 375 + 17625 + 33125 + 38500 + 38250 = 127875$$

In mod analog s-au determinat elementele de pe coloana 11 a tabelului 14.8. Se observă că indiferent de ordinea însumării, cele două sume duble sînt egale:

$$\sum_i \sum_j x_i y_j n_{ij} = \sum_j \sum_i y_j x_i n_{ij} \qquad \begin{matrix} i = 1,2, \ldots, k \\ j = 1,2, \ldots, p \end{matrix}$$

sau

$$\sum_i \sum_j x_i y_j n_{ij} = \sum_i y_j \sum_j x_i n_{ij}.$$

$$\sum_i \sum_j x_i y_j n_{ij} = \sum_i x_i \sum_j y_j n_{ij}$$

deci $$\sum_i x_i \sum_j y_j n_{ij} = \sum_j y_j \sum_i x_i n_{ij}$$

In cazul ecuaţiei de regresie $y_x = a+b\,x_i$ parametrii a şi b se calculează pe baza relaţiilor (14.32) şi (14.33):

$$a = \frac{27175.4525 - 905.127875}{39.27175 - 905^2} = \frac{289600}{240800} = 30,066$$

$$b = \frac{39.127875 - 905.4525}{39.27175 - 905^2} = \frac{891999.99}{240800} = 3,7043$$

Deci legătura dintre distribuţia personalului pe judeţe şi distribuţia fondurilor fixe este descrisă de ecuaţia :

$$y_x = 30,066 + 3,7043\,x_i$$

Parametrii ecuaţiei $x_y = a' + b'y_i$ se determină pe baza relaţiilor (14.35) şi (14.36), astfel:

$$a' = \frac{905.466875 - 4525 \cdot 127875}{39.466875 - 4525^2} = 68,849$$

$$b' = \frac{39 \cdot 127875 - 905 \cdot 4525}{39 \cdot 466875 - 4525^2} = -0,393$$

Legătura dintre distribuţia fondurilor fixe pe judeţe şi distribuţia numărului personalului este de forma:

$$x_y = 68,849 - 0,393\,y_i$$

Ecuaţia dreptei x_y sau y_x nu este o ecuaţie generală a oricărei linii drepte care descrie legătura dintre cele două variabile, ci reprezintă cîte o ecuaţie particulară a dreptei faţă de care suma pătratelor abaterilor valorilor empirice x_i , respectiv y_i , este minimă. Dacă abaterile în jurul acestor drepte sînt normal distribuite, aceste drepte vor fi cele mai potrivite pentru regresie. Dar dreapta de regresie obţinută pe baza metodei celor mai mici pătrate nu este în mod necesar cea mai bună ecuaţie de regresie , deoarece:

1. S-ar putea ca legătura dintre variabile să nu fie liniară, în care caz ecuaţia liniară de regresie nu este cea mai potri-

vită ecuaţie de regresie, ci doar cea mai bună dintre ecuaţiile de formă liniară;

2. Mai mult decît atît, nu este în mod necesar cea mai bună dreaptă de regresie. Ea are doar o şansă mai bună de a fi mai potrivită decît oricare altă dreaptă;

3. Dacă abaterile $y_i - y_{x_i}$ sau $x_i - x_{y_i}$ nu sînt normal distribuite, atunci nu avem temei să considerăm că dreapta obţinută pe baza metodei celor mai mici pătrate este mai bună decît cele obţinute pe baza altor metode.

In practica statistică metoda celor mai mici pătrate este cea mai des utilizată pentru estimarea parametrilor ecuaţiilor de regresie deoarece este simplă, obiectivă şi indicatorii se regresie se pot interpreta probabilistic.

14.2. Intensitatea legăturii liniare

14.2.1. Coeficientul de corelaţie

Determinarea ecuaţiei liniare de regresie în fiecare caz particular oferă imaginea formei legăturii dintre două variabile, dar nu răspunde la întrebarea, foarte importantă şi anume cît de puternică sau cît de intensă este legătura din variabilele corelate. Pentru a rezolva această problemă, să considerăm o distribuţie bidimensională situată în cadranul întîi al axelor de coordonate rectangulare (fig.14.6).

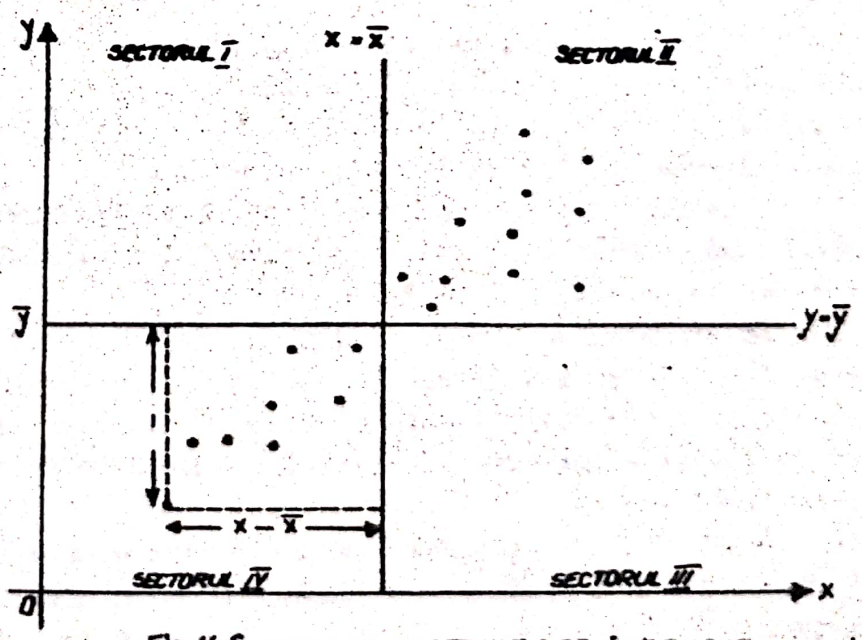


Fig.14.6. MEDIILE ŞI ABATERILE A DOUĂ VARIABILE CORELATE

Dreptele $x = \bar{x}$ şi $y = \bar{y}$ corespunzătoare mediilor variabilelor x şi y împart cadranul întîi în patru sectoare, în care abaterile variantelor individuale ale celor două variabile faţă de mediile lor pot fi pozitive sau negative. Astfel:

- în sectorul I abaterile variantelor variabilei X faţă de media lor $\bar{x}$ sînt negative, iar cele corespunzătoare variabilei Y sînt pozitive: $(x_i - \bar{x}) < 0$ şi $(y_i - \bar{y}) > 0$

- în sectorul II abaterile ambelor variabile faţă de mediile lor sînt pozitive:

$$(x_i - \bar{x}) > 0 \quad \text{şi} \quad (y_i - \bar{y}) > 0$$

- în sectorul III abaterile variabilei X sînt pozitive iar cele ale variabilei Y sînt negative:

$$(x_i - \bar{x}) > 0 \quad \text{şi} \quad (y_i - \bar{y}) < 0$$

- în sectorul IV abaterile ambelor variabile sînt negative:

$$(x_i - \bar{x}) < 0 \quad \text{şi} \quad (y_i - \bar{y}) < 0$$

Considerînd acum produsele abaterilor celor două variabile şi suma lor pentru fiecare sector, observăm că aceste sume sînt pozitive în sectoarele II şi IV şi negative în sectoarele I şi III.

Cînd majoritatea sau chiar totalitatea punctelor corespunzătoare perechilor de valori corelate x_i, y_i se situează în sectoarele II şi IV atunci înseamnă că între variabilele X şi Y există o corelaţie liniară directă şi puternică.

Cînd majoritatea sau chiar totalitatea punctelor se situează în sectoarele I şi III, corelaţia dintre variabile va fi puternică, inversă şi de formă liniară.

In sfîrşit, dacă o parte din puncte se situează în sectoarele II şi IV şi o altă parte în I şi III, aceasta sugerează existenţa unei distribuţii întîmplătoare a punctelor în cadranul I al axelor de coordonate, ceea ce înseamnă o legătură slabă între variabile.

Intensitatea legăturii liniare dintre două variabile se măsoară cu ajutorul coeficientului de corelaţie simplă liniară -

notat cu r - şi se calculează ca o medie artimetică simplă a produsului abaterilor normate ale celor două variabile.

Dacă notăm abaterile normate:

$$z_x = \frac{x_i - \bar{x}}{\sigma_x} \quad \text{şi} \quad z_y = \frac{y_i - \bar{y}}{\sigma_y}$$

atunci

$$r_{xy} = r_{yx} = \overline{z_x z_y} = \frac{\sum \frac{x_i - \bar{x}}{\sigma_x} \frac{y_i - \bar{y}}{\sigma_y}}{n} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n \sigma_x \sigma_y} \qquad (14.37)$$

Coeficientul de corelaţie este o mărime abstractă, independentă de unităţile de măsură a celor două variabile, fapt care-i măreşte posibilităţile de comparabilitate.

Coeficientul de corelaţie r variază între -1 şi +1 :

$$-1 \leqslant r \leqslant +1 \qquad (14.38)$$

Cînd r = +1 ecuaţiile de regresie y_x şi x_y coincid, ceea ce înseamnă că legătura dintre variabile este de tip funcţional, adică fiecărei valori date ale lui x îi corespunde o valoare pentru y şi numai una. In acest caz ambele variabile variază în acelaşi sens, iar legătura dintre ele poartă denumirea de <u>corelaţie pozitivă (directă) perfectă</u>.

In cazul opus, cînd r = -1, perechile de valori corelate x_i, y_i se înscriu tot pe o dreaptă ca şi în cazul anterior, cu deosebirea că valorilor mari ale lui x le corespund valori mici ale lui y şi invers. Legătura dintre variabile este tot de tip funcţional şi poartă denumirea de corelaţie <u>negativă (inversă) perfectă</u>.

Dacă variabilele sînt independente r = 0 deoarece, după cum vom vedea în continuare, coeficienţii de regresie ai celor două drepte $y_x = a_1 + b_1 x$ şi $x_y = a_2 + b_2 x$, sînt egali cu zero, ceea ce înseamnă că dreptele de regresie sînt paralele cu axele de coordonate. Reciproca nu este întotdeauna valabilă. Dacă r = 0 nu înseamnă că cele două variabile sînt independente în general, ci sînt numai necorelate liniar. Iată un exemplu care sugerează mai bine această idee. Să considerăm că perechile de valori x_i, y_i se prezintă ca în fig. 14.7.

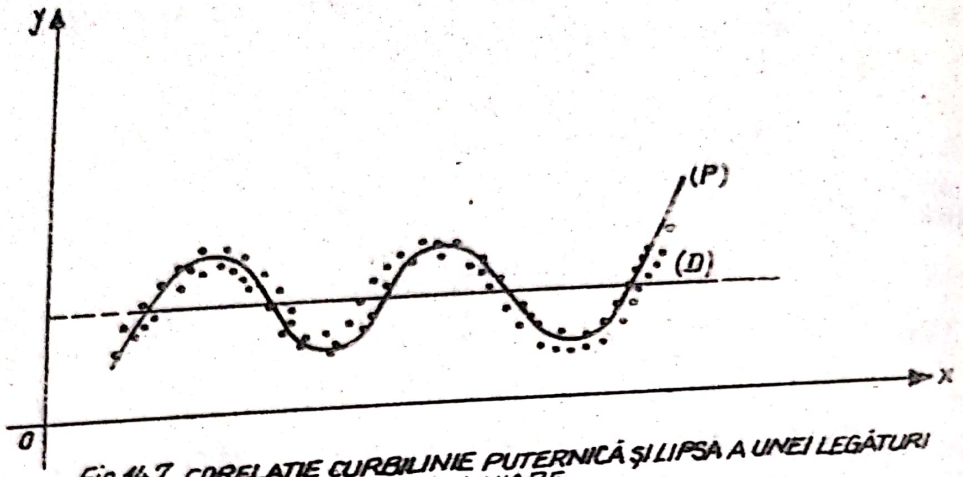
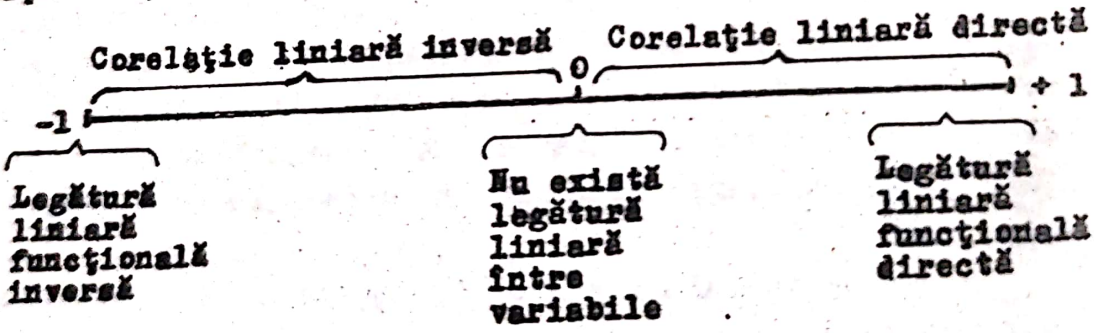


Fig.14.7. CORELAȚIE CURBILINIE PUTERNICĂ ȘI LIPSA A UNEI LEGĂTURI LINIARE

In acest caz, cea mai bună dreaptă de regresie este dreapta D paralelă cu axa Ox, deci cu coeficientul de regresie nul și r = 0. Dacă se renunță la modelul liniar de regresie și se consideră curba sinusoidală P, se observă că aceasta reprezintă o legătură sinusoidală sau ciclică perfectă. Ar fi deci greșit să considerăm că dacă r = 0 nu există nici un fel de legătură între variabile.

Interpretarea diferitelor valori ale coeficientului de corelație se poate prezenta cu ajutorul schemei următoare:



Cu cît r se apropie mai mult de cele două valori extreme, cu atît intensitatea legăturii este mai puternică și invers. Dacă r = 0 aceasta înseamnă că între variabile nu există o legătură liniară, dar poate să existe o legătură curbilinie. La interpretarea coeficientului de corelație, ca de altfel a oricărui indicator al intensității corelației, trebuie să se țină seama de numărul perechilor de valori corelate.

Pentru o mai bună stabilitate şi pentru a nu depinde prea mult de fluctuaţiile de selecţie, indicatorii de regresie şi corelaţie trebuie să se calculeze pentru un număr suficient de mare de valori ale varabilelor ce se corelează. Aceasta înseamnă că numărul de valori ale variabilelor trebuie să mai mare sau cel puţin egal cu 30. Astfel, dacă numărul de valori ale variabilelor luate în calcul este foarte mare, un coeficient de corelaţie foarte mic, să zicem 0,3 sau 0,4, va arăta desigur o legătură slabă dar stabilă, reală. Dacă nu dispunem decît de un număr mic de perechi de valori, să zicem 10, atunci un coeficient de corelaţie ce se apropie de unitate nu va avea decît o semnificaţie îndoielnică.

Să considerăm pentru exemplificarea calculului coeficientului de corelaţie liniară exemplul din tabelul 14.1, referitor la legătura dintre producţia de energie termoelectrică şi producţia de cărbune extras în România în perioada 1960-1976.

Schematic, calculele intermediare pentru determinarea coeficientului de corelaţie se prezintă în tabelul 14.9.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{309,2}{17} = 18,188 \text{ milioane tone}$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{442,6}{17} = 26,035 \text{ miliarde kwh}$$

$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{960,318}{17}} = 7,5159 \text{ milioane tone cărbune}$$

$$\sigma_x = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}} = \sqrt{\frac{3034,855}{17}} = 13,361 \text{ miliarde kwh.}$$

Coeficientul de corelaţie liniară se determină cu ajutorul formulei 14.37 :

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n \quad \sigma_x \quad \sigma_y} = \frac{1687,251}{17.7,5159.13,361} = 0,98$$

r = 0,98 arată că între producţia de cărbune şi producţia de energie termoelectrică există o corelaţie directă foarte intensă.

Scanned with OKEN Scanner

Tabelul 14.9. Calculele intermediare necesare pentru determinarea lui r

Anul	Cărbune extras (mil.to.) x_i	Energie termo - electr. (miliarde kwh) y_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
0	1	2	3	4	5	6	7
1960	8,2	7,3	- 9,988	99,760	-18,735	351,000	187,125
1961	8,7	8,2	- 9,488	90,022	-17,835	318,087	169,218
1962	9,6	9,4	- 8,588	73,754	-16,635	276,723	153,167
1963	10,3	11,1	- 7,888	62,221	-14,935	223,054	117,807
1964	11,1	13,3	- 7,088	50,240	-12,735	162,180	90,266
1965	12,1	16,2	- 6,088	37,064	- 9,835	96,727	59,875
1966	13,5	19,8	- 4,688	21,977	- 6,235	38,875	29,230
1967	15,0	23,3	- 3,188	10,163	- 2,735	7,480	8,719
1968	17,0	26,3	- 1,188	1,411	0,265	0,070	0,315
1969	19,2	29,3	1,012	1,024	3,265	10,660	3,304
1970	22,8	32,3	4,612	21,271	6,265	39,250	28,894
1971	23,0	35,0	4,812	23,155	8,965	80,371	43,140
1972	25,3	36,1	7,112	50,581	10,065	101,304	71,582
1973	26,7	39,2	8,512	72,454	13,165	173,317	112,060
1974	29,2	40,6	11,012	121,264	14,565	212,139	160,390
1975	29,4	45,0	11,212	125,709	18,965	359,671	212,636
1976	28,1	50,2	9,912	98,248	24,165	583,947	239,523
Total	309,2	442,6	-	960,318	-	3034,855	1687,251

Scanned with OKEN Scanner

Formula (14.37) de calcul a coeficientului de corelaţie se poate utiliza în practică în cazul în care variantele celor două variabile X şi Y se prezintă sub forma a două serii paralele. Determinarea acestui indicator poate fi efecuată mai uşor pe baza unei formule de calcul simplificat. Pentru stabilirea formulei de calcul simplificat se poate porni de la relaţia (14.37).

Substituind pe $\bar{x}$, $\bar{y}$, σ_x şi σ_y cu expresiile lor dezvoltate şi apoi efectuînd simplificările posibile se obţine:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n\,\sigma_x\,\sigma_y} = \frac{\sum(x_i y_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y})}{n\sqrt{\frac{\sum(x_i-\bar{x})^2}{n}\cdot\frac{\sum(y_i-\bar{y})^2}{n}}} =$$

$$= \frac{\sum x_i y_i - \sum x_i\bar{y} - \sum \bar{x}y_i + \sum \bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - 2\sum x_i\bar{x} + \sum \bar{x}^2)(\sum y_i^2 - 2\sum y_i\bar{y} + \sum \bar{y}^2)}} =$$

$$= \frac{\sum x_i y_i - \sum x_i \frac{\sum y_i}{n} - \frac{\sum x_i}{n}\sum y_i + \frac{\sum x_i}{n}\cdot\frac{\sum y_i}{n}\cdot n}{\sqrt{(\sum x_i^2 - 2\frac{\sum x_i}{n}\sum x_i + n\frac{\sum x_i}{n}\frac{\sum x_i}{n})(\sum y_i^2 - 2\frac{\sum y_i}{n}\sum y_i + n\frac{\sum y_i}{n}\cdot\frac{\sum y_i}{n})}}$$

de unde rezultă:

$$r = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[n\sum x_i^2 - (\sum x_i)^2][n\sum y_i^2 - (\sum y_i)^2]}} \qquad (14.39)$$

Calculul coeficientului de corelaţie simplă liniară este mult simplificat dacă se utilizează formula (14.39) deoarece toate sumele ce intervin în formulă sînt deja determinate cu ocazia calculului parametrilor ecuaţiei de regresie (tabelul 14.3).

$$r = \frac{17.972646 - 309{,}2 \cdot 442{,}6}{\sqrt{(17.6584{,}12 - 309{,}2^2)(17.14558{,}08 - 442{,}6^2)}} = \frac{28497{,}9}{29021{,}877} = 0{,}98$$

Se observă că s-a obţinut acelaşi rezultat ca şi în cazul utilizării formulei 14.37, deoarece formulele 14.37 şi 14.39 sînt echivalente.

Scanned with OKEN Scanner

Dacă perechile de valori x_i, y_i ale celor două variabile empirice sînt ponderate cu frecvenţa n_i , atunci formula coeficientului de corelaţie (14.37) devine:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})\, n_i}{(\sum n_i)\ \sigma_x\ \sigma_y} \qquad (14.40)$$

în care $\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2 n_i}{\sum n_i}}$ şi $\sigma_y = \sqrt{\frac{\sum (y_i - \bar{y})^2 n_i}{\sum n_i}}$

Efectuînd calculele în relaţia (14.40) rezultă formula de calcul simplificat a coeficientului de corelaţie:

$$r = \frac{\sum n_i \sum x_i y_i n_i - \sum x_i n_i \sum y_i n_i}{\sqrt{\left[\sum n_i \sum x_i^2 n_i - (\sum x_i n_i)^2\right]\left[\sum n_i \sum y_i^2 n_i - (\sum y_i n_i)^2\right]}} \qquad (14.41)$$

In cazul prezentării datelor sub forma unui tabel de corelaţie, formula coeficientului de corelaţie (14.37) devine:

$$r = \frac{\sum \sum (x_i - \bar{x})\ (y_j - \bar{y})\ n_{ij}}{(\sum \sum n_{ij})\ \sigma_x\ \sigma_y} \qquad (14.42)$$

în care:

$$\sigma_y = \sqrt{\frac{\sum (x_i - \bar{x})^2 n_{i.}}{\sum n_{i.}}} \quad \text{şi} \quad \sigma_y = \sqrt{\frac{\sum (y_j - \bar{y})^2 n_{.j}}{\sum n_{.j}}}$$

Pe baza relaţiei (14.42) se obţine următoarea formulă de calcul simplificat a coeficientului de corelaţie:

$$r = \frac{\sum \sum n_{ij} \sum \sum x_i y_j n_{ij} - \sum x_i n_{i.} \sum y_j n_{.j}}{\sqrt{\left[\sum \sum n_{ij} \sum x_i^2 n_i - (\sum x_i n_i)^2\right]\left[\sum \sum n_{ij} \sum y_j^2 n_{.j} - (\sum y_j n_{.j})^2\right]}} \qquad (14.43)$$

Folosirea formulei (14.43) pentru calculul coeficientului de corelaţie pe baza datelor din tabelul 14.7 necesită întocmirea

în prealabil a tabelului 14.1o.

Pe baza datelor din tabelul 14.8 se pot determina mediile celor două variabile:

$$\bar{x} = \frac{\sum x_i n_{i\cdot}}{\sum n_{i\cdot}} = \frac{905}{39} = 23,205$$

și

$$\bar{y} = \frac{\sum y_j n_{\cdot j}}{\sum n_{\cdot j}} = \frac{4525}{39} = 116,026$$

Substituind datele din tabelul 141ó în formula (14.42) rezultă:

$$r = \frac{22871,794}{39 \cdot 12,582 \cdot 55,322} = 0,84$$

Coeficientul de corelație $r = 0,84$ evidențiază faptul că între numărul mediu al personalului și valoarea fondurilor fixe există o corelație directă și intensă.

O simplificare mai mare a calculelor se va obține dacă se substituie valorile x_i și y_i ale variabilelor cu x_i', respectiv y_i', date de următoarele relații:

$$x_i' = \frac{x_i - x_0}{k_x} \quad \text{și} \quad y_i' = \frac{y_i - y_0}{k_y} \qquad (14.44)$$

unde x_0, respectiv y_0 reprezintă nivelurile din centrul celor două distribuții, iar k_x și k_y reprezintă cel mai mare divizor comun al diferențelor față de originea de calcul aleasă.

Substituind (14.44) în relațiile (14.39), (14.41) și (14.43) se obțin alte formule de calcul simplificat ale coeficientului de corelație, astfel:

- în cazul cînd perechile de valori x_i, y_i sînt neponderate:

$$r = \frac{n\sum \frac{x_i - x_0}{k_x} \cdot \frac{y_i - y_0}{k_y} - \sum \frac{x_i - x_0}{k_x} \cdot \sum \frac{y_i - y_0}{k_y}}{\sqrt{\left[n\sum \left(\frac{x_i - x_0}{k_x}\right)^2 - \left(\sum \frac{x_i - x_0}{k_x}\right)^2\right]\left[n\sum \left(\frac{y_i - y_0}{k_y}\right)^2 - \left(\sum \frac{y_i - y_0}{k_y}\right)^2\right]}} \qquad (14.45)$$

- în cazul ponderării perechilor de valori corelate x_i, y_i cu n_i :

$$r = \frac{\sum n_i \sum \frac{x_i - x_0}{k_x} \cdot \frac{y_i - y_0}{k_y} \cdot n_i - \sum \frac{x_i - x_0}{k_x} \cdot n_i \cdot \sum \frac{y_i - y_0}{k_y} \cdot n_i}{\sqrt{\left[\sum n_i \sum \left(\frac{x_i - x_0}{k_x}\right)^2 n_i - \left(\sum \frac{x_i - x_0}{k_n} n_i\right)^2\right]\left[\sum n_i \sum \left(\frac{y_i - y_0}{k_y}\right)^2 n_i - \left(\sum \frac{y_i - y_0}{k_y} n_i\right)^2\right]}} \qquad (14.46)$$

- în cazul unui tabel de corelaţie:

$$r = \frac{\sum\sum n_{ij} \sum\sum \frac{x_i - x_0}{k_x} \cdot \frac{y_j - y_0}{k_y} n_{ij} - \sum \frac{x_i - x_0}{k_x} n_{i\cdot} \cdot \sum \frac{y_j - y_0}{k_y} n_{\cdot j}}{\sqrt{\left[\sum\sum n_{ij} \sum \left(\frac{x_i - x_0}{k_x}\right)^2 n_{i\cdot} - \left(\sum \frac{x_i - x_0}{k_x} n_{i\cdot}\right)^2\right]\left[\sum\sum n_{ij} \sum \left(\frac{y_j - y_0}{k_y}\right)^2 n_{\cdot j} - \left(\sum \frac{y_j - y_0}{k_y} n_{\cdot j}\right)^2\right]}} \qquad (14.47)$$

Să exemplificăm utilizarea formulei (14.47) pe baza datelor din tabelul 14.7. Sistematizarea calculelor preliminare necesare determinării coeficientului de corelaţie în acest caz se prezintă în tabelul 14.11.

Considerînd: $x_0 = x_3 = 25$ şi $y_0 = y_3 = 125$ iar ca divizori comuni ai diferenţelor $x_i - x_0$ şi $y_j - y_0$ vor fi $k_x = 10$, iar $k_y = 50$.

Substituind elementele de calcul din tabelul 14.11 în formula 14.47 se obţine :

$$r = \frac{39.47 - (-7)(-7)}{\sqrt{\left[39.63 - (-7)^2\right]\left[39.49 - (-7)^2\right]}} = \frac{1784}{2117,4739} = 0,84$$

S-a obţinut aceeaşi valoare a coeficientului de corelaţie ca şi în cazul aplicării formulei (14.42), dar cu un volum de muncă mult mai redus. Astfel, pentru efectuarea calculelor din tabelul 14.11 nu este necesar să se apeleze la nici un mijloc modern de calcul, deoarece se lucrează cu numere mici şi întregi, iar calculele se pot face mintal.

- în cazul ponderării perechilor de valori corelate x_i; y_i cu n_i :

$$r = \frac{\sum n_i \sum \frac{x_i - x_0}{k_x} \cdot \frac{y_i - y_0}{k_y} \cdot n_i - \sum \frac{x_i - x_0}{k_x} \cdot n_i \cdot \sum \frac{y_i - y_0}{k_y} \cdot n_i}{\sqrt{\left[\sum n_i \sum \left(\frac{x_i - x_0}{k_x}\right)^2 n_i - \left(\sum \frac{x_i - x_0}{k_n} n_i\right)^2\right]\left[\sum n_i \sum \left(\frac{y_i - y_0}{k_y}\right)^2 n_i - \left(\sum \frac{y_i - y_0}{k_y} n_i\right)^2\right]}} \qquad (14.46)$$

- în cazul unui tabel de corelaţie:

$$r = \frac{\sum\sum n_{ij} \sum\sum \frac{x_i - x_0}{k_x} \cdot \frac{y_i - y_0}{k_y} n_{ij} - \sum \frac{x_i - x_0}{k_x} n_{i\cdot} \cdot \sum \frac{y_i - y_0}{k_y} n_{\cdot j}}{\sqrt{\left[\sum\sum n_{ij} \sum \left(\frac{x_i - x_0}{k_x}\right)^2 n_{i\cdot} - \left(\sum \frac{x_i - x_0}{k_x} n_{i\cdot}\right)^2\right]\left[\sum\sum n_{ij} \sum \left(\frac{y_i - y_0}{k_y}\right)^2 n_{\cdot j} - \left(\sum \frac{y_i - y_0}{k_y} n_{\cdot j}\right)^2\right]}} \qquad (14.47)$$

Să exemplificăm utilizarea formulei (14.47) pe baza datelor din tabelul 14.7. Sistematizarea calculelor preliminare necesare determinării coeficientului de corelaţie în acest caz se prezintă în tabelul 14.11.

Considerînd: $x_0 = x_3 = 25$ şi $y_0 = y_3 = 125$ iar ca divizori comuni ai diferenţelor $x_i - x_0$ şi $y_j - y_0$ vor fi $k_x = 10$, iar $k_y = 50$.

Substituind elementele de calcul din tabelul 14.11 în formula 14.47 se obţine :

$$r = \frac{39.47 - (-7)(-7)}{\sqrt{\left[39.63 - (-7)^2\right]\left[39.49 - (-7)^2\right]}} = \frac{1784}{2117,4739} = 0,84$$

S-a obţinut aceeaşi valoare a coeficientului de corelaţie ca şi în cazul aplicării formulei (14.42), dar cu un volum de muncă mult mai redus. Astfel, pentru efectuarea calculelor din tabelul 14.11 nu este necesar să se apeleze la nici un mijloc modern de calcul, deoarece se lucrează cu numere mici şi întregi, iar calculele se pot face mintal.

Tabelul 14.10. Calculele preliminare necesare pentru determinarea lui r prin metoda momentelor

y / x	-50	50-100	100-150	150-200	200 -	$n_{i\cdot}$	x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^2 n_{i\cdot}$	$\sum_j (x_i-\bar{x})(y_j-\bar{y})n$
1	2	3	4	5	6	7	8	9	10	11	12
- 10	3	3	-	-	-	6	5	-18,205	331,422	1988,532	7212,020
10 - 20	-	8	3	-	-	11	15	-8,205	67,322	740,542	2472,052
20 - 30	-	4	6	2	-	12	25	1,795	3,222	38,664	13,800
30 - 40	-	-	2	1	1	4	35	11,795	139,122	556,488	2192,643
40 -	-	-	-	3	3	6	45	21,795	475,022	2850,132	10981,279
$n_{\cdot j}$	3	15	11	6	4	39				6174,358	22871,794
y_j	25	75	125	175	225	-					
$y_j-\bar{y}$	-91,026	-41,026	8,974	58,974	108,974	-					
$(y_j-\bar{y})^2$	8285,733	1683,1326	80,533	3477,933	11875,332	-					
$(y_j-\bar{y})^2 n_{\cdot j}$	24857,197	25246,989	885,859	20867,595	47501,328	119358,968					
$\sum (x_i-\bar{x})(y_j-\bar{y})n_{ij}$	4971,385	4639,015	87,452	4763,330	8410,612	22871,794					

Scanned with OKEN Scanner

14.3.1. Raportul de corelaţie

Dacă coeficientul de corelaţie este un indicator al intensităţii legăturii numai în cazul celor liniare, cunoscînd valorile empirice ale variabilei rezultative y_i şi valorile teoretice corespunzătoare y_x se poate stabili un indicator general valabil al intensităţii legăturii dintre două fenomene, numit <u>raport de corelaţie</u> pe care-l vom nota în continuare cu η . Calculul lui η se bazează pe descompunerea dispersiei totale a variabilei empirice Y pe factori de influenţă.

Variaţia variabilei empirice Y depinde de mărimea influenţei tuturor factorilor cauzali care acţionează asupra ei, inclusiv valoarea influenţei factorului X. Variaţia variabilei teoretice Y_x depinde numai de influenţa factorului X. Aceasta înseamnă că dispersia variabilei empirice Y se poate descompune în dispersia valorilor empirice faţă de valorile teoretice Y_x şi dispersia valorilor teoretice faţă de media lor. Sugestiv aceste abateri se prezintă în fig. 14.8.

Abaterile valorilor empirice faţă de media lor se pot descompune în două componente şi anume abaterea valorilor teoretice faţă de media lor şi abaterea valorilor empirice faţă de valorile lor teoretice:

$$y_i - \bar{y} = (y_i - y_x) + (y_x - \bar{y})$$

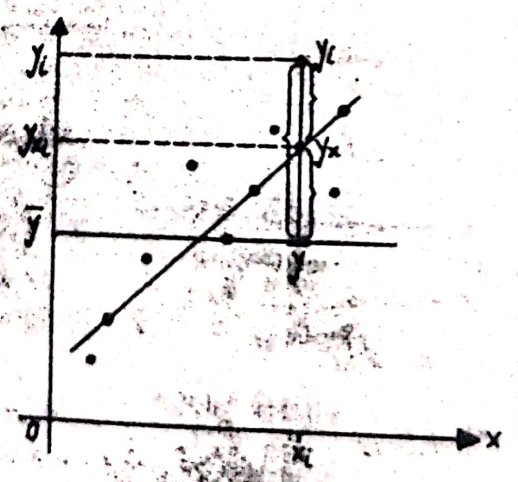


Fig. 14.8. DESCOMPUNEREA VARIAŢIEI TOTALE ÎN VARIAŢIE EXPLICATĂ ŞI VARIAŢIE

Ridicînd la pătrat abaterile şi însumînd, rezultă:

$$\sum(y_i - \bar{y})^2 = \sum(y_i - y_x)^2 + \sum(y_x - \bar{y})^2$$

Impărţind ambii membri ai acestei relaţii prin numărul perechilor de valori corelate "n", se obţin următoarele dispersii:

- dispersia generală a variabilei Y :

$$\sigma_y^2 = \frac{\sum(y_i - \bar{y})^2}{n}$$

- dispersia valorilor teoretice ale variabilei faţă de media lor:

$$\sigma_{y_x}^2 = \frac{\sum(y_{x_i} - \bar{y})^2}{n}$$

- dispersia valorilor empirice faţă de valorile teoretice ale variabilei Y , numită şi dispersie reziduală:

$$\sigma_{yy_x}^2 = \frac{\sum(y_i - y_x)^2}{n}$$

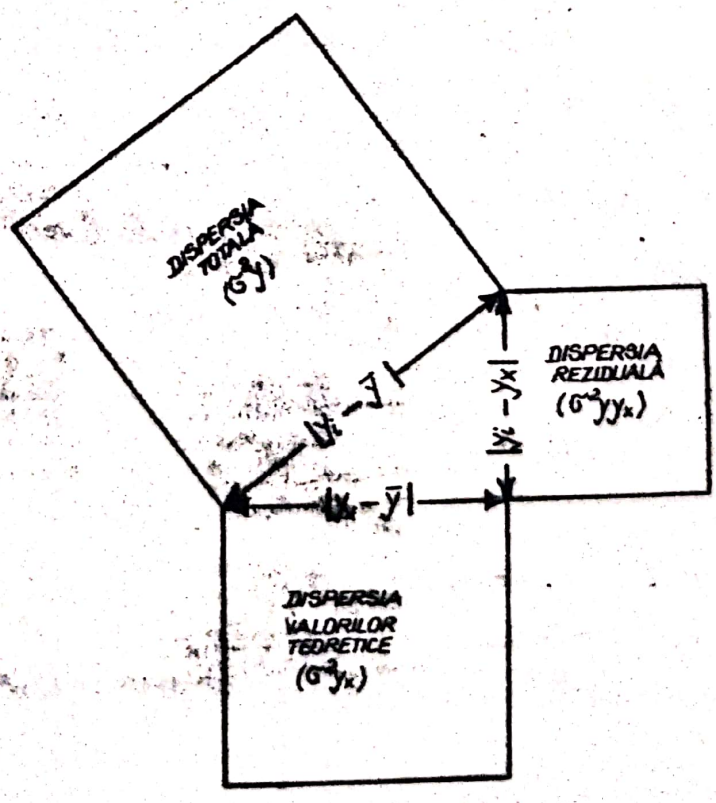


FIG. 4.9. RELAŢIA DINTRE DISPERSIA VALORILOR EMPIRICE ALE LUI Y COMPONENTELE EI FACTORIALE

Tabelul 14.11. Calculul simplificat al lui r în cazul unei distribuţii bidimensionale

Nr. ort.	x \ y	-50	50-100	100-150	150-200	200-	$n_{i\cdot}$	x_i	x_i-x_o	$\frac{x_i-x_o}{k_x}$	$\frac{x_i-x_o}{k_x}n_{i\cdot}$	$\frac{(x_i-x_o)^2}{k_x}n_{i\cdot}$	$\sum_i \frac{x_i-x_o}{k_x}\frac{y_j-y_o}{k_y}n_{ij}$
o	1	2	3	4	5	6	7	8	9	10	11	12	13
1	- 10	3	3	-	-	-	6	5	-20	-2	-12	24	18
2	10 - 20	-	8	3	-	-	11	15	-10	-1	-11	11	8
3	20 - 30	-	4	6	2	-	12	25	0	0	0	0	0
4	30 - 40	-	-	2	1	1	4	35	10	1	4	4	3
5	40 -	-	-	-	3	3	6	45	20	2	12	24	18
6	$n_{\cdot j}$	3	15	11	6	4	39	-	-	-	-7	63	47
7	y_j	25	75	125	175	225	-						
8	y_j-y_o	-100	-50	0	50	100	-						
9	$\frac{y_j-y_o}{k_y}$	-2	-1	0	1	2	-						
10	$\frac{y_j-y_o}{k_y}n_{\cdot j}$	-6	-15	0	6	8	-7						
11	$\frac{(y_j-y_o)^2}{k_y}n_{\cdot j}$	12	15	0	6	16	49						
12	$\sum \frac{x_i-x_o}{k_x}\cdot\frac{y_j-y_o}{k_y}n_{ij}$	12	14	0	7	14	47						

Scanned with OKEN Scanner

Dispersia generală σ_y^2 reflectă influenţa tuturor factorilor care au acţionat asupra variabilei dependente Y, inclusiv influenţa variabilei independente X, în timp ce dispersia valorilor teoretice y_{x_i} faţă de media lor $\bar{y}$ măsoară numai influenţa variabilei independente X asupra variabilei dependente, iar influenţa celorlalţi factori care au acţionat asupra variabilei Y, cu excepţia variabilei X, este dată de expresia reziduală. Aceasta înseamnă că între aceste trei categorii de dispersii există următoarea relaţie:

$$\sigma_y^2 = \sigma_{y_x}^2 + \sigma_{yy_x}^2 \qquad (14.48)$$

In cazul unor legături funcţionale între variabilele X şi Y, întreaga variaţie a variabilei dependente se datorează influenţei variabilei independente X şi, prin urmare, dispersia reziduală este egală cu zero. In această situaţie valorile empirice y_i coincid cu valorile teoretice y_x.

$$y_i = y_x \quad \text{şi} \quad \sigma_y^2 = \sigma_{y_x}^2 \quad \text{iar} \quad \sigma_{yy_x}^2 = 0$$

Gradul de dependenţă dintre variabilele X şi Y sau intensitatea legăturii dintre ele se poate aprecia pe baza raportului dintre dispersia valorilor teoretice şi dispersia valorilor empirice ale variabilei Y.

Acest raport (variaţiei) măsoară deci ponderea influenţei factorului X asupra variabilei Y faţă de variaţia lui Y pe seama tuturor factorilor de influenţă inclusiv X. Indicatorul, astfel obţinut, poartă denumirea de <u>coeficient de determinaţie</u>, se notează cu η_d^2 şi se calculează pe baza relaţiei:

$$\eta_d^2 = \frac{\sigma_{y_x}^2}{\sigma_y^2} \qquad (14.49)$$

Ţinînd seama de relaţia (14.48), coeficientul de determinaţie devine:

$$\eta_d^2 = \frac{\sigma_y^2 - \sigma_{yy_x}^2}{\sigma_y^2} = 1 - \frac{\sigma_{yy_x}^2}{\sigma_y^2} \qquad (14.50)$$

Din relaţia (14.50) rezultă că, coeficientul de determinaţie η_d^2 variază între zero şi 1 :

$$0 \leq \eta_d^2 \leq 1$$

$\eta_d^2 = 1$ atunci cînd întreaga variaţie a vraibilei Y este determinată de variaţia variabilei independente X , deci cînd dispersia reziduală este egală cu zero.

$\eta_d^2 = 0$ atunci cînd variabilele X şi Y sînt independente şi deci dispersia reziduală este egală cu dispersia generală a variabilei Y .

Raportînd dispersia reziduală la dispersia totală a variabilei Y se obţine <u>coeficientul de nedeterminaţie</u> (η_n^2) :

$$\eta_n^2 = \frac{\sigma_{yy_x}^2}{\sigma_y^2} \qquad (14.51)$$

şi

$$\eta_n^2 = \frac{\sigma_y^2 - \sigma_{y_x}^2}{\sigma_y^2} = 1 - \frac{\sigma_{y_x}^2}{\sigma_y^2} = 1 - \eta_d^2$$

deci:

$$\eta_d^2 + \eta_n^2 = 1 \qquad (14.52)$$

şi

$$\eta_d^2 = 1 - \eta_n^2 \qquad (14.53)$$

$$\eta_n^2 = 1 - \eta_d^2 \qquad (14.54)$$

Indicatorul intensităţii corelaţiei de acelaşi grad cu valorile empirice se obţine prin extragerea rădăcinii pătrate din coeficientul de determinaţie şi poartă denumirea de <u>raport de corelaţie</u>(η) :

$$\eta = \sqrt{\frac{\sigma_{y_x}^2}{\sigma_y^2}} \qquad (14.55)$$

sau

$$\eta = \sqrt{1 - \frac{\sigma_{yy_x}^2}{\sigma_y^2}} \qquad (14.56)$$

Substituind în formula (14.55) pe $\sigma^2_{y_x}$ şi σ^2_y cu expresiile lor se obţine:

$$\eta = \sqrt{\frac{\frac{\sum(y_x - \bar{y})^2}{n}}{\frac{\sum(y_i - \bar{y})^2}{n}}} = \sqrt{\frac{\sum(y_x - \bar{y})^2}{\sum(y_i - \bar{y})^2}} \qquad (14.57)$$

Efectuînd aceeaşi substituţie în (14.56) rezultă:

$$\eta = \sqrt{1 - \frac{\sum(y_i - y_x)^2}{\sum(y_i - \bar{y})^2}} \qquad (14.58)$$

Din formula (14.56) rezultă că raportul de corelaţie variază în intervalul (o,1) : $0 \leq \eta \leq 1$

Calculul şi analiza raportului de corelaţie se vor efectua pe acelaşi exemplu pentru care s-a calculat coeficientul de corelaţie referitor la legătura dintre numărul mediu al personalului şi valoarea fondurilor fixe pe judeţe în anul 1976 (tabelul 14.1).

In paragraful 14.1 s-a obţinut ecuaţia de regresie a celor două variabile:

$$y_x = -5,7144 + 1,7456\, x_i$$

iar valorile teoretice ale variabilei Y s-au determinat în coloana 6 din tabelul 14.3.

Calcule preliminare necesare determinării raportului de corelaţie simplă liniară se prezintă, în mod sistematizat, în tabelul 14.12.

Substituind în formula (14.58) elementele de calcul din tabelul 14.12 rezultă:

$$\eta = \sqrt{1 - \frac{108,392}{3034,855}} = \sqrt{0,9643} = 0,98$$

Se observă că raportul de corelaţie $\eta = 0,98$ are aceeaşi valoare ca şi coeficientul de corelaţie calculat pe baza aceloraşi date. Coeficientul de determinaţie $\eta^2_d = 96,43\%$ evi-

denţiază faptul că 96,43% din variaţia producţiei de energie termoelectrică se datorează variaţiei producţiei de cărbune şi numai 3,57% altor factori.

Tabelul 14.12. Elementele de calcul preliminar ale lui η

Anul	x_i	y_i	y_x	$(y_i-y_x)^2$	$(y_i-\bar{y})^2$
0	1	2	3	4	5
1960	8,2	7,3	8,599	1,687	351,000
1961	8,7	8,2	9,472	1,618	318,087
1962	9,6	9,4	11,043	2,699	276,723
1963	10,3	11,1	12,265	1,357	223,054
1964	11,1	13,3	13,662	0,131	162,180
1965	12,1	16,2	15,407	0,629	96,727
1966	13,5	19,8	17,851	3,799	38,875
1967	15,0	23,3	20,470	8,009	7,480
1968	17,0	26,3	23,961	5,471	0,070
1969	19,2	29,3	27,801	2,247	10,660
1970	22,8	32,3	34,085	3,186	39,250
1971	23,0	35,0	34,434	0,320	80,371
1972	25,3	36,1	38,449	5,518	101,304
1973	26,7	39,2	40,893	2,866	173,317
1974	29,2	40,6	45,257	21,688	212,139
1975	29,4	45,0	45,606	0,066	359,671
1976	28,1	50,2	43,337	47,101	583,947
Total	309,2	442,6	442,592	108,392	3034,855

In cazul unei legături simple liniare raportul de corelaţie se determină pe baza formulei (14.58). Pentru explicitarea expresiilor celor două dispersii $\sigma^2_{yy_x}$ şi σ^2_y se poate porni de la relaţia care există între valorile empirice şi valorile teoretice ale variabilei Y :

$$y_i - y_x = y_i - a - b\,x_i \qquad (14.59)$$

Înmulţind cu $y - y_x$ ambii membri ai relaţiei (14.59) şi însumînd, rezultă:

$$\sum(y_i-y_x)^2 = \sum(y_i-a-b\,x_i)(y_i-y_x) = \sum(y_i-y_x)y_i - a\sum(y_i-y_x) - b\sum(y_i-y_x)x_i$$

denţiază faptul că 96,43% din variaţia producţiei de energie termoelectrică se datorează variaţiei producţiei de cărbune şi numai 3,57% altor factori.

Tabelul 14.12. Elementele de calcul preliminar ale lui η

Anul	x_i	y_i	y_x	$(y_i-y_x)^2$	$(y_i-\bar{y})^2$
0	1	2	3	4	5
1960	8,2	7,3	8,599	1,687	351,000
1961	8,7	8,2	9,472	1,618	318,087
1962	9,6	9,4	11,043	2,699	276,723
1963	10,3	11,1	12,265	1,357	223,054
1964	11,1	13,3	13,662	0,131	162,180
1965	12,1	16,2	15,407	0,629	96,727
1966	13,5	19,8	17,851	3,799	38,875
1967	15,0	23,3	20,470	8,009	7,480
1968	17,0	26,3	23,961	5,471	0,070
1969	19,2	29,3	27,801	2,247	10,660
1970	22,8	32,3	34,085	3,186	39,250
1971	23,0	35,0	34,434	0,320	80,371
1972	25,3	36,1	38,449	5,518	101,304
1973	26,7	39,2	40,893	2,866	173,317
1974	29,2	40,6	45,257	21,688	212,139
1975	29,4	45,0	45,606	0,066	359,671
1976	28,1	50,2	43,337	47,101	583,947
Total	309,2	442,6	442,592	108,392	3034,855

In cazul unei legături simple liniare raportul de corelaţie se determină pe baza formulei (14.58). Pentru explicitarea expresiilor celor două dispersii $\sigma^2_{yy_x}$ şi σ^2_y se poate porni de la relaţia care există între valorile empirice şi valorile teoretice ale variabilei Y :

$$y_i - y_x = y_i - a - b\,x_i \qquad (14.59)$$

Inmulţind cu $y - y_x$ ambii membri ai relaţiei (14.59) şi insumînd, rezultă:

$$\sum(y_i-y_x)^2 = \sum(y_i-a-b\,x_i)(y_i-y_x) = \sum(y_i-y_x)y_i - a\sum(y_i-y_x) - b\sum(y_i-y_x)x_i$$

Dar: $\sum(y_i-y_x) = 0$ deoarece reprezintă tocmai derivata a întîia a sumei de pătrate (S) în raport cu a

şi

$\sum(y_i-y_x)x_i = 0$ fiind derivata a întîia a sumei de pătrate (S) în raport cu b .

Deci $$\sum(y_i-y_x)^2 = \sum(y_i-y_x)y_i \qquad (14.60)$$

Înmulţinid ambii membri al ecuaţiei (14.60) cu y_i şi însumînd se obţine:

$$\sum(y_i-y_x)y_i = \sum(y_i-a-b\,x_i)y_i$$

Deci $$\sum(y_i-y_x)^2 = \sum y_i^2 - a\sum y_i - b\sum x_i y_i \qquad (14.61)$$

Iar expresia de la numitorul raportului de corelaţie devine prin transformări elementare:

$$\sum(y_i-\bar{y})^2 = \sum(y_i^2-2\,y_i\bar{y}+\bar{y}^2) = \sum y_i^2 - 2\,\frac{(\sum y_i)^2}{n} + n\,\frac{(\sum y_i)^2}{n^2}$$

Deci
$$\sum(y_i-\bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n} \qquad (14.62)$$

Substituind în formula raportului de corelaţie (14.58) relaţiile (14.61) şi (14.62), rezultă:

$$\eta = \sqrt{1 - \frac{\sum(y_i-y_x)^2}{\sum(y_i-\bar{y})^2}} = \sqrt{1 - \frac{\sum y_i^2 - a\sum y_i - b\sum x_i y_i}{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}} =$$

$$= \sqrt{\frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n} - \sum y_i^2 + a\sum y_i + b\sum x_i y_i}{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}}$$

Deci
$$\eta = \sqrt{\frac{a\sum y_i + b\sum x_i y_i - \frac{(\sum y_i)^2}{n}}{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}} \qquad (14.63)$$

Scanned with OKEN Scanner

Calculul simplificat al raportului de corelaţie simplă liniară se exemplifică pe baza datelor din tabelul 14.3.

Substituind în formula (14.63) elementele de calcul din tabelul 14.3 obţinem:

$$\eta = \sqrt{\frac{(-5,7144).442,6+1,7456.9726,46 - \frac{442,6^2}{17}}{14558,08 - \frac{442,6^2}{17}}} = \sqrt{0,9642} = 0,98$$

S-a obţinut acelaşi rezultat ca şi în cazul aplicării formulei de bază a raportului de corelaţie (14.58).

Dacă perechile de valori corelate sînt ponderate cu n_i, atunci formula raportului de corelaţie (14.63) devine:

$$\eta = \sqrt{\frac{a\sum y_i n_i + b\sum x_i y_i n_i - \frac{(\sum y_i n_i)^2}{\sum n_i}}{\sum y_i^2 n_i - \frac{(\sum y_i n_i)^2}{\sum n_i}}} \qquad (14.64)$$

In cazul unor distribuţii bidimensionale (x_i, y_j, n_{ij}) se aplică o formulă modificată, care să ţină seama de indicii de variaţie ai celor două variabile şi de frecvenţa perechilor de valori corelate n_{ij} :

$$\eta = \sqrt{\frac{a\sum y_j n_{\cdot j} + b\sum\sum x_i y_j n_{ij} - \frac{(\sum y_j n_{\cdot j})^2}{\sum\sum n_{ij}}}{\sum y_j^2 n_{\cdot j} - \frac{(\sum y_j n_{\cdot j})^2}{\sum\sum n_{ij}}}} \qquad (14.65)$$

Coeficientul de corelaţie simplă liniară (r) este egal cu rădăcina pătrată din coeficientul de determinaţie (η_d^2) :

$$\eta_d^2 = r^2 \qquad (14.66)$$

Pentru demonstrarea egalităţii (14.66) pornim de la relaţia (14.55) în care, după transformări elementare, numărăto-

rul devine:

$$\sum(y_x-\bar{y})^2 = n\bar{y}^2+\sum y_x^2-2\bar{y}\sum y_x = n\bar{y}^2 + \sum(a+bx)^2-2\bar{y}\sum(a+bx) =$$

$$= \frac{(\sum y)^2}{n} + na^2 + b^2\sum x^2 + 2ab\sum x - 2a\sum y - \frac{2b\sum x\sum y}{n}$$

Substituind $a = \frac{\sum y - b\sum x}{n}$ rezultă:

$$\sum(y_x-\bar{y})^2 = \frac{(\sum y)^2}{n} + n\frac{(\sum y-b\sum x)^2}{n} + b^2\sum x^2 + 2b\frac{\sum y-b\sum x}{n}\sum x -$$

$$- 2\frac{\sum y-b\sum x}{n}\sum y - \frac{2b\sum x\sum y}{n}$$

şi

$$n\sum(y_x-\bar{y})^2 = (\sum y)^2+(\sum y-b\sum x)^2+ nb^2\sum x^2+2b\sum x\sum y-2b^2(\sum x)^2 -$$

$$- 2(\sum y)^2 + 2b\sum x\sum y - 2b\sum x\sum y =$$

$$= (\sum y)^2 +(\sum y)^2+b^2(\sum x)^2-2b\sum x\sum y + nb^2\sum x^2 +$$

$$+ 2b\sum x\sum y - 2b^2(\sum x)^2 - 2(\sum y)^2 =$$

$$= nb^2\sum x^2 - b^2(\sum x)^2 = b^2[n\sum x^2-(\sum x)^2]$$

Efectuînd calculele în expresia de la numitorul lui η_d^2 se obţine:

$$\sum(y-\bar{y})^2 = \sum y^2 - 2\bar{y}\sum y + n\bar{y}^2 = \sum y^2 - 2\frac{(\sum y)^2}{n} + \frac{(\sum y)^2}{n}$$

şi

$$n\sum(y-\bar{y})^2 = \sum y^2 - \frac{(\sum y)^2}{n} = n\sum y^2 - (\sum y)^2$$

Deci

$$\eta_d^2 = \frac{b^2[n\sum x^2-(\sum x)^2]}{n\sum y^2-(\sum y)^2}$$

Considerînd $b = \frac{n\sum xy - \sum x\sum y}{n\sum x^2 - (\sum x)^2}$ se obţine:

Scanned with OKEN Scanner

$$\eta_d^2 = \frac{\left[\frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}\right]^2 \left[n\sum x^2 - (\sum x)^2\right]}{n\sum y^2 - (\sum y)^2} =$$

$$= \frac{(n\sum xy - \sum x \sum y)^2}{\left[n\sum x^2 - (\sum x)^2\right]\left[n\sum y^2 - (\sum y)^2\right]} = r^2$$

Deci relaţia (14.66) este demonstrată.

14.3. Legătura dintre coeficienţii de regresie şi coeficientul de corelaţie

In cazul regresiei lui Y în funcţie de X, între coeficientul de regresie b_{yx} şi coeficientul de corelaţie r_{yx} există următoarea relaţie:

$$b_{yx} = r_{yx} \frac{\sigma_y}{\sigma_x} \qquad (14.67)$$

de unde:

$$r_{yx} = b_{yx} \frac{\sigma_x}{\sigma_y} \qquad (14.68)$$

Pentru verificarea relaţiei (14.68) considerăm:

$$b_{yx} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$\sigma_x = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n}}$$

$$\sigma_y = \sqrt{\frac{\sum y^2 - \frac{(\sum y)^2}{n}}{n}}$$

Substituind aceste elemente în relaţia (14.68) rezultă:

Scanned with OKEN Scanner

$$r_{yx} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} \cdot \frac{\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}}{\sqrt{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}} =$$

$$= \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} \cdot \frac{\sqrt{n\Sigma x^2 - (\Sigma x)^2}}{\sqrt{n\Sigma y^2 - (\Sigma y)^2}} = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

In mod analog, în cazul regresiei lui x în funcţie de y se verifică relaţiile:

$$b_{xy} = r_{xy} \frac{\sigma_x}{\sigma_y} \qquad (14.69)$$

$$r_{xy} = b_{xy} \frac{\sigma_y}{\sigma_x} \qquad (14.70)$$

Din relaţiile (14.69) şi (14.70) rezultă că un coeficient de corelaţie poate fi egal cu coeficientul de regresie corespunzător atunci cînd dispersiile celor două variabile corelate sînt egale.

$$r_{yx} = b_{yx} \quad \text{cînd} \quad \sigma_x = \sigma_y \qquad (14.71)$$

$$r_{xy} = b_{xy} = b_{yx} \qquad (14.72)$$

Între coeficienţii de regresie b_{yx}, b_{xy} şi coeficientul de corelaţie r există următoarea relaţie:

$$r_{xy} = \sqrt{b_{yx} \cdot b_{xy}} \qquad (14.73)$$

Pentru verificarea acestei relaţii să substituim în (14.73) coeficienţii de regresie cu expresiile lor:

$$r_{xy} = \sqrt{b_{xy} \cdot b_{yx}} =$$

$$= \sqrt{\frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} \cdot \frac{n\Sigma yx - \Sigma y \Sigma x}{n\Sigma y^2 - (\Sigma y)^2}} = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

In particular, dacă cei doi coeficienţi de regresie sînt mărimi reciproce unul faţă de celălalt, cele două drepte de regresie coincid, iar coeficientul de corelaţie r va fi egal cu 1, deci corelaţia este perfectă, funcţională.

Relaţia (14.73 ne permite să determinăm unul dintre cele trei elemente pe baza cunoaşterii a două dintre ele. Astfel, în mod obişnuit se cunoaşte b_{xy} şi r şi pe baza acestora se poate determina b_{yx}.

15. REGRESIE SI CORELATIE SIMPLA CURBILINIE

15.1. Modele de regresie simplă curbilinie

Alături de modelele liniare de regresie, un loc important în stabilirea expresiei analitice a legăturilor dintre fenomene îl ocupă modelele unor curbe, de diferite tipuri: hiperbolice, parabolice, exponenţiale, logaritmice etc. Legăturile liniare, exprimate sintetic prin ecuaţia unei drepte, se caracterizează prin faptul că unei variaţii uniforme a variabilei factoriale îi corespunde o variaţie tot uniformă a variabilei rezultative. Pentru orice nivel al avrabilei independente la o creştere sau scădere a acesteia cu o unitate, îi corespunde o creştere sau descreştere a variabilei dependente Y cu aceeaşi mărime, egală cu coeficientul de regresie b. Legăturile de tip liniar sînt cele mai simple modele de regresie şi se întîlnesc destul de rar în cadrul fenomenelor reale din natură şi societate.

In practică cele mai frecvente legături statistice dintre fenomenele reale iau forma unor modele curbilinii, în care variaţia variabilei rezultative depinde de însuşi nivelul la care se află variabila independentă. Teoretic, se poate imagina un număr nelimitat de curbe care pot fi descrise de funcţii matematice, în realitate însă se utilizează un număr redus de modele de regresie curbilinie şi anume:

1. Modele parabolice:

- parabola de gradul doi $y = a+b\,x + c\,x^2$ (15.1)
- parabola de gradul trei $y = a+bx + cx^2+dx^3$ (15.2)
- parabola semilogaritmică de gradul doi

$$\log x = a + bx + cx^2 \qquad (15.3)$$

sau

$$y = a + b \log x + c(\log x)^2 \qquad (15.4)$$

Scanned with OKEN Scanner

- parabola semilogaritmică de gradul trei:

$$\log y = a + bx + cx^2 + dx^3 \qquad (15.5)$$

$$\text{sau} \qquad y = a + b \log x + c(\log x)^2 + d(\log x)^3 \qquad (15.6)$$

- parabola dublu logaritmică de gradul doi:

$$\log y = a + b \log x + c(\log x)^2 \qquad (15.7)$$

- parabola dublu logaritmică de gradul trei:

$$\log y = a + b \log x + c(\log x)^2 + d(\log x)^3 \qquad (15.8)$$

2. Modele hiperbolice:

$$y = a + \frac{b}{x} \qquad (15.9)$$

$$y = \frac{1}{a + bx} \qquad \text{sau} \qquad \frac{1}{y} = a + bx \qquad (15.10)$$

$$y = \frac{x}{a\,x + b} \qquad (15.11)$$

3. Modele de tip exponenţial:

$$y = a\, e^{bx} \qquad (15.12)$$

$$y = a\, e^{\frac{b}{x}} \qquad (15.13)$$

$$y = \frac{1}{a + b\, e^{-x}} \qquad (15.14)$$

4. Modelul unei funcţii putere:

$$y = a\, x^b \qquad (15.15)$$

5. Modelul unei funcţii de tip logaritmic:

$$y = a + b \log x \qquad (15.16)$$

15.1.1. Legături de tip parabolic

Parabola de gradul doi $y = a + bx + cx^2$, a cărei expresie grafică se prezintă în fig. 15.1, are o serie de particularităţi a căror cunoaştere poate să orienteze mai bine statisticia-

nul în alegerea acestui tip de funcţie drept model de regresie.

Aşa după cum se observă din fig. 15.1 şi fig. 15.2 parabola de gradul doi obişnuită sau semilogaritmică prezintă un punct de extrem, de maxim sau de minim, în funcţie de semnul coeficientului de regresie c. Dacă $c < 0$ atunci curba parabolei de gradul doi este convexă, iar dacă $c > 0$ curba este concavă. Cunoaşterea abscisei punctului de extrem este deosebit de importantă în activitatea practică, deoarece reprezintă nivelul variabilei factoriale care corespunde nivelului maxim sau celui minim al variabilei rezultative.

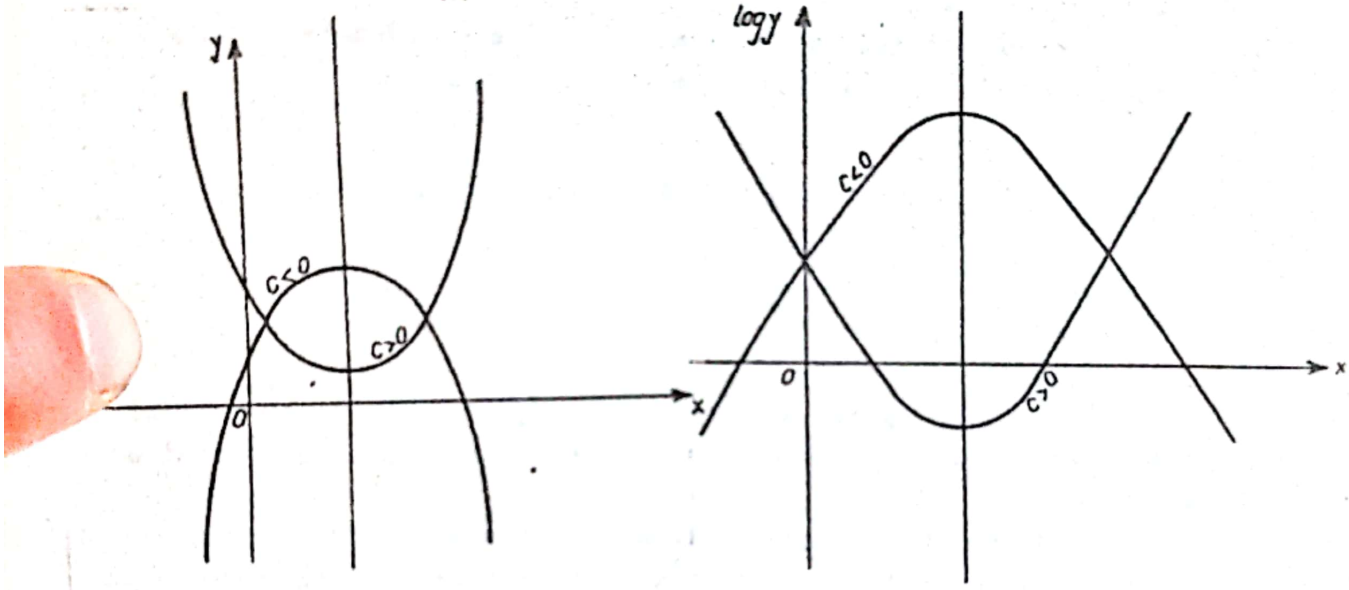


Fig. 15.1. PARABOLA DE GRADUL DOI $y = a + bx + cx^2$ Fig. 15.2. PARABOLA SEMILOGARITMICĂ DE GRADUL DOI $\log y = a + bx + cx^2$

Abscisa punctului de extrem al parabolei de gradul doi $y = a + b\,x + cx^2$ se obţine, după cum se ştie, egalînd cu zero derivata de ordinul întîi, astfel:

$$y' = b + 2\,c\,x = 0 \qquad (15.17)$$

de unde

$$x = -\frac{b}{2\,c} \qquad (15.18)$$

Desigur în practica statistică, legăturile de tip parabolic se vor prezenta sub forma unor segmente ale curbei crescătoare sau descrescătoare, situate de regulă în cadranul întîi al axelor de coordonate, deoarece indicatorii sociali şi economici reali au valori pozitive.

Indiferent de forma ei, concavă sau convexă, parabola de gradul doi este o curbă simetrică faţă de ordonata punctului de extrem.

Ca şi în cazul dreptei de regresie, parametrii parabolei de gradul doi se pot determina cu ajutorul metodei punctelor selecţionate sau cu ajutorul metodei celor mai mici pătrate.

Intrucît parabola de gradul doi depinde de trei parametri, se vor selecţiona trei puncte echidistante şi se va obţine un sistem format din trei ecuaţii. Punctele selecţionate se vor situa cîte unul la fiecare capăt al segmentului parabolei şi unul la mijlocul ei.

Dacă punctele selecţionate sînt de coordonate (x_1, y_1), (x_2, y_2), (x_3, y_3) atunci se poate scrie sistemul:

$$a + b\,x_1 + c\,x_1^2 = y_1$$
$$a + b\,x_2 + c\,x_2^2 = y_2$$
$$a + b\,x_3 + c\,x_3^2 = y_3$$

Rezolvînd acest sistem se vor obţine valorile celor trei necunoscute care sînt tocmai parametrii a, b şi c ai parabolei de gradul doi.

Pentru determinarea parametrilor ecuaţiei de regresie de tip parabolic se utilizează în mod obişnuit, metoda celor mai mici pătrate, ceea ce înseamnă că se determină minimul expresiei:

$$S = \sum(y_i - a - b\,x_i - c\,x_i^2)^2 = \text{minim} \qquad (15.19)$$

Calculînd derivatele parţiale ale funcţiei S în raport cu a , b şi c , şi egalîndu-le cu zero, rezultă:

$$\frac{\partial S}{\partial a} = 2\sum(y_i - a - b\,x_i - c\,x_i^2)(-1) = 0$$
$$\frac{\partial S}{\partial b} = 2\sum(y_i - a - b\,x_i - c\,x_i^2)(-x_i) = 0 \qquad (15.20)$$
$$\frac{\partial S}{\partial c} = 2\sum(y_i - a - b\,x_i - c\,x_i^2)(-x_i^2) = 0$$

Prin transformări elementare sistemul (15.20) devine:

$$\begin{aligned} na + b\sum x_i + c\sum x_i^2 &= \sum y_i \\ a\sum x_i + b\sum x_i^2 + c\sum x_i^3 &= \sum x_i y_i \\ a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 &= \sum x_i^2 y_i \end{aligned} \qquad (15.21)$$

Pentru rezolvarea sistemului (15.21) este necesar să se determine în prealabil expresiile $\sum x_i$, $\sum x_i^2$, $\sum x_i^3$, $\sum x_i^4$, $\sum y_i$, $\sum x_i y_i$ și $\sum x_i^2 y_i$. Sistematizarea acestor calcule se prezintă în tabelul 15.1.

Tabelul 15.1. Calculele preliminare pentru determinarea parametrilor unei parabole de gradul doi

Nr. crt.	y_i	x_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
1	y_1	x_1	x_1^2	x_1^3	x_1^4	$x_1 x_1$	$x_1^2 y_1$
2	y_2	x_2	x_2^2	x_2^3	x_2^4	$x_2 y_2$	$x_2^2 y_2$
3	y_3	x_3	x_3^2	x_3^3	x_3^4	$x_3 y_3$	$x_3^2 y_3$
⋮							
i	y_i	x_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
⋮							
n	y_n	x_n	x_n^2	x_n^3	x_n^4	$x_n y_n$	$x_n^2 y_n$
TOTAL	$\sum y_i$	$\sum x_i$	$\sum x_i^2$	$\sum x_i^3$	$\sum x_i^4$	$\sum x_i y_i$	$\sum x^2 y_i$

Dacă frecvența perechilor de valori corelate x_i, y_i este diferită de unitate, atunci sistemul (15.21) devine:

$$\begin{aligned} a\sum n_i + b\sum x_i n_i + c\sum x_i^2 n_i &= \sum y_i n_i \\ a\sum x_i n_i + b\sum x_i^2 n_i + c\sum x_i^3 n_i &= \sum x_i y_i n_i \\ a\sum x_i^2 n_i + b\sum x_i^3 n_i + c\sum x_i^4 n_i &= \sum x_i^2 y_i n_i \end{aligned} \qquad (15.22)$$

In acest caz, calculul sumelor pentru cele două variabile, ce apar în sistem, se prezintă sistematizat în tabelul 15.2.

Dacă datele inițiale se prezintă sub forma unui tabel de corelație, sistemul (15.21) devine:

$$\begin{aligned} a\sum\sum n_{ij} + b\sum x_i n_{i\cdot} + c\sum x_i^2 n_{i\cdot} &= \sum y_j n_{\cdot j} \\ a\sum x_i n_{i\cdot} + b\sum x_i^2 n_{i\cdot} + c\sum x_i^3 n_{i\cdot} &= \sum\sum x_i y_j n_{ij} \\ a\sum x_i^2 n_{i\cdot} + b\sum x_i^3 n_{i\cdot} + c\sum x_i^4 n_{i\cdot} &= \sum\sum x_i^2 y_j n_{ij} \end{aligned} \qquad (15.23)$$

Tabelul 15.2

Nr. crt.	y_i	x_i	n_i	$y_i n_i$	$x_i n_i$	$x_i^2 n_i$	$x_i^3 n_i$	$x_i^4 n_i$	$x_i y_i n_i$	$x_i^2 y_i n_i$
1	y_1	x_1	n_1	$y_1 n_1$	$x_1 n_1$	$x_1^2 n_1$	$x_1^3 n_1$	$x_1^4 n_1$	$x_1 y_1 n_1$	$x_1^2 y_1 n_1$
2	y_2	x_2	n_2	$y_2 n_2$	$x_2 n_2$	$x_2^2 n_2$	$x_2^3 n_2$	$x_2^4 n_2$	$x_2 y_2 n_2$	$x_2^2 y_2 n_2$
3	y_3	x_3	n_3	$y_3 x_3$	$x_3 n_3$	$x_3^2 n_3$	$x_3^3 n_3$	$x_3^4 n_3$	$x_3 y_3 n_3$	$x_3^2 y_3 n_3$
...										
i	y_i	x_i	n_i	$y_i n_i$	$x_i n_i$	$x_i^2 n_i$	$x_i^3 n_i$	$x_i^4 n_i$	$x_i y_i n_i$	$x_i^2 y_3 n_3$
...										
k	y_k	x_k	n_k	$y_k x_k$	$x_k n_k$	$x_k^2 n_k$	$x_k^3 n_k$	$x_k^4 n_k$	$x_k y_k n_k$	$x_k^2 y_k n_k$
Total	-	-	$\sum n_i$	$\sum y_i n_i$	$\sum x_i n_i$	$\sum x_i^2 n_i$	$\sum x_i^3 n_i$	$\sum x_i^4 n_i$	$\sum x_i y_i n_i$	$\sum x_i^2 y_i n_i$

Calculul sumelor simple și duble ce apar în sistemul (15.23) se sistematizează cu ajutorul tabelul 15.3.

Sumele duble ce apar în tabelul 15.3. se mai pot calcula și astfel:

$$\sum_i x_i y_j n_{ij} = y_j \sum_i x_i n_{ij}$$

$$\sum_j x_i y_j n_{ij} = x_i \sum_j y_j n_{ij}$$

iar

Tabel 15.3

Nr. crt.	Y \ X	x_1	x_2	...	x_i	...	x_s	Total $n_{\cdot j}$	$y_j n_{\cdot j}$	$\sum_i y_j x_i n_{ij}$	$\sum_i y_j x_i^2 n_{ij}$	$y_j^2 n_{\cdot j}$
0	A	1	2	...	i	...	s	s+1	s+2	s+3	s+4	s+5
1	y_1	n_{11}	n_{21}	...	n_{i1}	...	n_{s1}	$n_{\cdot 1}$	$y_1 n_{\cdot 1}$	$\sum y_1 x_i n_{i1}$	$\sum y_1 x_i^2 n_{i1}$	$y_1^2 n_{\cdot 1}$
2	y_2	n_{12}	n_{22}	...	n_{i2}	...	n_{s2}	$n_{\cdot 2}$	$y_2 n_{\cdot 2}$	$\sum y_2 x_i n_{i2}$	$\sum y_2 x_i^2 n_{i2}$	$y_2^2 n_{\cdot 2}$
:												
j	y_j	n_{1j}	n_{2j}	...	n_{ij}	...	n_{sj}	$n_{\cdot j}$	$y_j n_{\cdot j}$	$\sum y_j x_i n_{ij}$	$\sum y_j x_i^2 n_{ij}$	$y_j^2 n_{\cdot j}$
:												
k	y_k	n_{1k}	n_{2k}	...	n_{ik}	...	n_{sk}	$n_{\cdot k}$	$y_k n_{\cdot k}$	$\sum y_k x_i n_{ik}$	$\sum y_k x_i^2 n_{ik}$	$y_k^2 n_{\cdot k}$
k+1	Total $n_{i\cdot}$	n_1	n_2	...	n_i	...	n_s	n	$\sum y_j n_{\cdot j}$	$\sum\sum y_j x_i n_{ij}$	$\sum\sum y_j x_i^2 n_{ij}$	$\sum y_j^2 n_{\cdot j}$
k+2	$x_i n_{i\cdot}$	$x_1 n_{1\cdot}$	$x_2 n_{2\cdot}$	...	$x_i n_{i\cdot}$	...	$x_s n_{s\cdot}$	$\sum x_i n_{i\cdot}$				
k+3	$x_i^2 n_{i\cdot}$	$x_1^2 n_{1\cdot}$	$x_2^2 n_{2\cdot}$	...	$x_i^2 n_{i\cdot}$	...	$x_s^2 n_{s\cdot}$	$\sum x_i^2 n_{i\cdot}$				
k+4	$x_i^3 n_{i\cdot}$	$x_1^3 n_{1\cdot}$	$x_2^3 n_{2\cdot}$	...	$x^3 n_{i\cdot}$	...	$x_s^3 n_{s\cdot}$	$\sum x_i^3 n_{i\cdot}$				
k+5	$x_i^4 n_{i\cdot}$	$x_1^4 n_{1\cdot}$	$x_2^4 n_{2\cdot}$	...	$x^4 n_{i\cdot}$	...	$x_s^4 n_{s\cdot}$	$\sum x_i^4 n_{i\cdot}$				
k+6	$\sum_j x_i y_j n_{ij}$	$\sum x_1 y_j n_{ij}$	$\sum x_2 y_j n_{2j}$	...	$\sum x_i y_j n_{ij}$	...	$\sum x_s y_j n_{sj}$	$\sum\sum x_i y_j n_{ij}$				
k+7	$\sum_j x_i^2 y_j n_{ij}$	$\sum x_1^2 y_j n_{1j}$	$\sum x_2^2 y_j n_{2j}$	...	$\sum x_i^2 y_j n_{ij}$	...	$\sum x_s^2 y_j n_{sj}$	$\sum\sum x_i^2 y_j n_{ij}$				
k+8	y_{x_i}	y_{x_1}	y_{x_2}	...	y_{x_i}	...	y_{x_s}	-				

Scanned with OKEN Scanner

$$\sum_i \sum_j x_i y_j n_{ij} = \sum_j (y_j \sum_i x_i n_{ij})$$

sau

$$\sum_i \sum_j x_i y_j n_{ij} = \sum_i (x_i \sum_j y_j n_{ij})$$

şi

$$\sum_i x_i^2 y_j n_{ij} = y_j \sum_i x_i^2 n_{ij}$$

$$\sum_j x_i^2 y_j n_{ij} = x_i^2 \sum_j y_j n_{ij}$$

de unde

$$\sum_i \sum_j x_i^2 y_j n_{ij} = \sum_j (y_j \sum_i x_i^2 n_{ij})$$

$$\sum_i \sum_j x_i^2 y_j n_{ij} = \sum_i (x_i^2 \sum_j y_j n_{ij})$$

Intensitatea corelaţiei simple parabolice se măsoară cu ajutorul raportului de corelaţie, definit de relaţia:

$$\eta = \sqrt{\frac{\sigma_y^2 - \sigma_{yy_x}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sum(y_i - y_x)^2}{\sum(y_i - \bar{y}^2}} \qquad (15.24)$$

Determinarea raportului de corelaţie cu ajutorul formulei (15.24), în cazul unei regresii după modelul unei parabole de gradul doi, presupune cunoaşterea valorilor teoretice ale variabilei rezultative y_x. Sistematizat, calculele preliminare, necesare determinării raportului de corelaţie, se prezintă în tabelul (15.4).

Calculul raportului de corelaţie simplă de tip parabolic se simplifică, în mare măsură, dacă în locul valorilor teoretice y_x vom introduce ecuaţia polinomului de gradul doi care s-ar obţine pe baza datelor iniţiale.

Tabelul 15.4

Nr. crt.	y_i	y_{x_i}	y_i-y_x	$(y_i-y_x)^2$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$
1	y_1	y_{x_1}	$y_1-y_{x_1}$	$(y_1-y_{x_1})^2$	$y_1-\bar{y}$	$(y_1-\bar{y})^2$
2	y_2	y_{x_2}	$y_2-y_{x_2}$	$(y_2-y_{x_2})^2$	$y_2-\bar{y}$	$(y_2-\bar{y})^2$
⋮						
i	y_i	y_{x_i}	$y_i-y_{x_i}$	$(y_i-y_{x_i})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$
⋮						
n	y_n	y_{x_n}	$y_n-y_{x_n}$	$(y_n-y_{x_n})^2$	$y_n-\bar{y}$	$(y_n-\bar{y})^2$
Total	$\sum y_i$	$\sum y_x$	-	$\sum(y_i-y_x)^2$	-	$\sum(y_i-y)^2$

Diferența dintre valorile empirice și cele teoretice ale variabilei rezultative se calculează astfel:

$$y_i - y_{x_i} = y_i - a - bx_i - cx_i^2 \qquad (15.25)$$

Inmulțind ambii membri ai ecuației (15.25) cu (y_i-y_x) și insumînd rezultă:

$$\sum(y_i-y_{x_i})^2 = \sum(y_i-y_{x_i})(y_i-a-bx_i-cx_i^2)$$

$$\sum(y_i-y_{x_i})^2 = \sum(y_i-y_{x_i})y_i - a\sum(y_i-y_{x_i}) - b\sum(y_i-y_{x_i})x_i - c\sum(y_i-y_{x_i})x_i^2 \qquad (15.26)$$

In membrul al doilea al acestei relații sumele din cadrul ultimilor trei termeni sînt nule, deoarece ele reprezintă cele trei derivate parțiale ale funcției S în raport cu parametrii a,b și c ai parabolei de gradul doi:

$$\frac{\partial S}{\partial a} = \sum(y_i-y_{x_i}) = \sum y_i - na - b\sum x_i - c\sum x_i^2 = 0$$

$$\frac{\partial S}{\partial b} = \sum(y_i-y_{x_i})x_i = \sum x_i y_i - a\sum x_i - b\sum x_i^2 - c\sum x_i^3 = 0 \qquad (15.27)$$

$$\frac{\partial S}{\partial c} = \sum(y_i-y_{x_i})x_i^2 = \sum x_i^2 y_i - a\sum x_i^2 - b\sum x_i^3 - c\sum x_i^4 = 0$$

In baza relaţiilor (15.27), ecuaţia (15.26) devine:

$$\sum(y_i-y_{x_i})^2 = \sum(y_i-y_{x_i})y_i$$

$$\sum(y_i-y_{x_i})y_i = \sum(y_i-a-bx_i-c\,x_i^2)y_i =$$

$$= \sum y_i^2 - a\sum y_i - b\sum x_iy_i - c\sum x_i^2y_i$$

Deci

$$\sum(y_i-y_{x_i})^2 = \sum y_i^2 - a\sum y_i - b\sum x_iy_i - c\sum x_i^2y_i \qquad (15.28)$$

Aceasta este expresia explicită a sumei pătrate din cadrul numărătorului formulei raportului de corelaţie (15.24).

Prin transformări elementare suma de pătrate de la numitorul lui η devine :

$$\sum(y_i-\bar{y})^2 = \sum y_i^2 - 2\sum y_i\,\frac{\sum y_i}{n} + \frac{n(\sum y_i)^2}{n^2}$$

$$\sum(y_i-\bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n} \qquad (15.29)$$

Substituind relaţiile (15.28) şi (15.29) în formula raportului de corelaţie (15.24), rezultă:

$$\eta = \sqrt{1 - \frac{\sum y_i^2 - a\sum y_i - b\sum x_iy_i - c\sum x_i^2y_i}{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}}$$

$$\eta = \sqrt{\frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n} - \sum y_i^2 + a\sum y_i - b\sum x_iy_i + c\sum x_i^2y_i}{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}}$$

$$= \sqrt{\frac{a\sum y_i + b\sum x_iy_i + c\sum x_i^2y_i - \frac{(\sum y_i)^2}{n}}{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}} \qquad (15.30)$$

Din formula (15.30) se observă că toate elementele de calcul ce intră în componenţa relaţiei au fost deja determinate cu ocazia estimării parametrilor a, b şi c ai parabolei.

Dacă perechile de valori empirice x_i, y_i ale celor două variabile corelate sînt ponderate cu frecvenţele n_i, atunci în formula raportului de corelaţie (15.30) va trebui ca fiecare variabilă să fie ponderată cu frecvenţa corespunzătoare:

$$\eta = \sqrt{\frac{a\sum y_i n_i + b\sum x_i y_i n_i + c\sum x_i^2 y_i n_i - \frac{(\sum y_i n_i)^2}{\sum n_i}}{\sum y_i^2 n_i - \frac{(\sum y_i n_i)^2}{\sum n_i}}} \qquad (15.31)$$

Atunci cînd datele iniţiale, cu privire la cele două variabile corelate, se prezintă sub forma unui tabel de corelaţie, formula raportului de corelaţie simplă parabolică devine:

$$\eta = \sqrt{\frac{a\sum y_j n_{\cdot j} + b\sum\sum y_j x_i n_{ij} + c\sum\sum x_i^2 y_j n_{ij} - \frac{(\sum y_j n_{\cdot j})^2}{\sum\sum n_{ij}}}{\sum y_j^2 n_{\cdot j} - \frac{(\sum y_j n_{\cdot j})^2}{\sum\sum n_{ij}}}} \qquad (15.32)$$

Să considerăm, pentru exemplificarea regresiei şi corelaţiei simple parabelice, legătura dintre valoarea fondurilor fixe şi valoarea producţiei globale industriale pe judeţele României în anul 1976. Datele cu privire la aceste două variabile se prezintă în tabelul 15.5.

Tabelul 15.5. Distribuţia judeţelor după valoarea fondurilor fixe şi valoarea producţiei globale în anul 1976 (miliarde lei)

Producţia globală (Y) / Fonduri fixe (X)	-10,0	10,1-20,0	20,1-30,0	30,1-40,0	40,1 -	Total judeţe
-10	6	-	-	-	-	6
10,1-20	7	3	-	-	-	10
20,1-30	4	7	2	-	-	13
30,1-40	-	2	2	-	-	4
40,1-	-	1	3	1	1	6
Total judeţe	17	13	7	1	1	39

Sursa: Prelucrat după Anuarul statistic al R.S.R. 1977, p.93,117.

S-a exclus din calcul Municipiul Bucureşti, deoarece acesta prezintă caracteristici specifice şi nu poate fi asimilat cu judeţele ţării.

Pentru a ne forma o primă imagine asupra formei legăturii dintre valoarea fondurilor fixe şi valoarea producţiei globale pe judeţe, vom reprezenta grafic această distribuţie bidimensională (fig. 15.3).

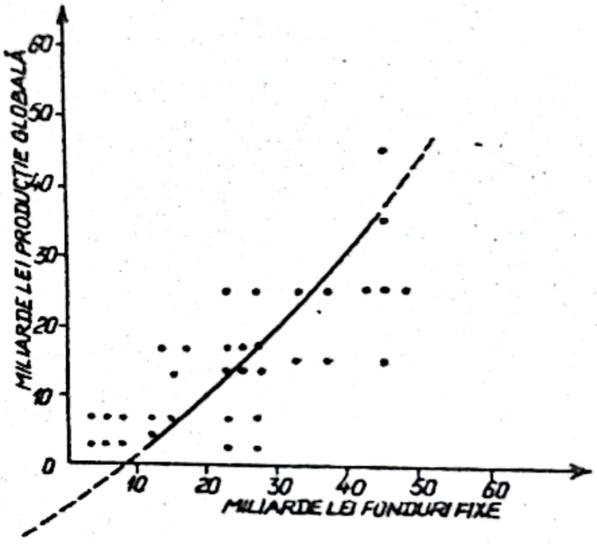


Fig 15.3 LEGĂTURA DINTRE VALOAREA FONDURILOR FIXE ŞI VALOAREA PRODUCŢIEI GLOBALE PE JUDEŢE ÎN ANUL 1976

Se observă din figura 15.3 că legătura dintre variabile se apropie de o parabolă de gradul doi.

Pentru determinarea parametrilor parabolei de gradul doi, pe baza datelor din tabelul 15.5, calculele preliminare se sistematizează cu ajutorul tabelului 15.6. Simplificarea calculelor se face considerînd valorile reduse x'_i şi y'_j ale celor două variabile obţinute prin schimbarea originii de calcul, astfel:

$$x'_i = \frac{x_i - x_0}{k_x} \quad \text{şi} \quad y'_j = \frac{y_j - y_0}{k_y}$$

Substituind elementele de calcul din tabelul 15.6 în sis-

Tabelul 15.6. Calculele preliminare pentru determinarea indicatorilor de regresie şi corelaţie simplă parabolică

X \ Y	-10,0	10,1-20,0	20,1-30,0	30,1-40,0	40,1-	Total ($n_{i\cdot}$)	x_i	$x_i n_{i\cdot}$	$x_i^2 n_{i\cdot}$	$x_i^3 n_{i\cdot}$	$\sum_j x_i y_j n_{ij}$	$\sum_j x_i^2 y_j n_{ij}$	$x_i^4 n_{i\cdot}$
1	2	3	4	5		7	8	9	10	11	12	13	14
-10,0	6	-	-	-	-	6	5	30	150	750	150	750	3750
10,1-20,0	7	3	-	-	-	10	15	150	2250	33750	1200	18000	506250
20,1-30,0	4	7	2	-	-	13	25	325	8125	203125	4375	109375	5078125
30,1-40,0	-	2	2	-	-	4	35	140	4900	171500	2800	98000	6002500
40,1-	-	1	3	1	1	6	45	270	12150	1761750	7650	344250	79278750
Total ($n_{\cdot j}$)	17	13	7	1	1	39	-	915	27575	2170875	16175	570375	90869375
y_j	5	15	25	35	45	-							
$y_j n_{\cdot j}$	85	195	175	35	45	535							
$y_j^2 n_{\cdot j}$	425	2925	4375	1225	2025	10975							
$\sum_i x_i y_j n_{ij}$	1175	5025	6375	1575	2025	16175							
$\sum_i x_i^2 y_j n_{ij}$	21125	142875	244375	70875	91125	570375							

Scanned with OKEN Scanner

temul (15.23), rezultă:

$$\begin{aligned} 39\,a + 915\,b + 27575\,c &= 535 \\ 915\,a + 27575\,b + 2170875\,c &= 16175 \\ 27575\,a + 2170875\,b + 90869375\,c &= 570375 \end{aligned}$$

Sistemul obţinut se poate rezolva prin metoda lui Gauss (tabelul 15.7).

Tabelul 15.7. Rezolvarea sistemului de ecuaţii normale prin metoda lui Gauss

	a	b	c	
a	39	915	27575	535
b	915	27575	2170875	16175
c	27575	2170875	90869375	570375
a	1	23,461568	707,051282	13,717948
b	0	6107,665280	1523922,2624	3623,0612
c	0	1523922,2624	71372435,8989	192102,5642
a	1	0	-5146,8478	- 0,1995
b	0	1	249,5098	0,5932
c	0	0	308961103,0073	-711886,711
a	1	0	0	-12,0372
b	0	1	0	1,1670
c	0	0	1	- 0,0023

Substituind parametrii a, b şi c , de pe ultima coloană a tabelului 15.7, în ecuaţia parabolei de grad doi (15.3), se obţine ecuaţia de regresie a producţiei globale industriale în funcţie de valoarea fondurilor fixe pe judeţe:

$$y_{x_i} = -12,0372 + 1,1670\,x_i - 0,0023\,x_i^2$$

Inlocuind succesiv valorile lui x_i în această ecuaţie rezultă:

$$\begin{aligned} x_1 &= 5 & y_1 &= -6,2597 \\ x_2 &= 15 & y_2 &= 4,9503 \\ x_3 &= 25 & y_3 &= 15,7003 \\ x_4 &= 35 & y_4 &= 24,1503 \\ x_5 &= 45 & y_5 &= 35,8203 \end{aligned}$$

Reprezentînd grafic valorile teoretice y_{x_i} în fig.15.3 se obţine un arc de parabolă ce trece prin centrul norului de puncte, ceea ce dovedeşte că ecuaţia de regresie de formă parabolică descrie în mod corespunzător legătura dintre valoarea producţiei globale şi valoarea fondurilor fixe pe judeţe.

Intensitatea legăturii dintre valoarea producţiei globale şi valoarea fondurilor fixe pe judeţe.

Intensitatea legăturii dintre valoarea producţiei globale industriale şi valoarea fondurilor fixe pe judeţe se poate determina cu ajutorul relaţiei (15.32), astfel:

$$\eta = \sqrt{\frac{-12,0372.535 + 1,167.16175 - 0,0023.570375 - \frac{535^2}{39}}{10975 - \frac{535^2}{39}}}$$

$$= \sqrt{0,8346} = 0,91$$

iar

$$\eta^2 = 83,46\%$$

Aceasta înseamnă că între gradul de înzestrare tehnică a judeţelor şi valoarea producţiei globale obţinute există o legătură parabolică intensă şi anume în medie 83,46% din valoarea producţiei globale industriale este determinată de valoarea fondurilor fixe pe judeţe şi numai 16,54% se datorează altor factori.

In cazul parabolei de gradul doi semilogaritmică (15.3), reprezentată grafic în fig. 15.2, prin aplicarea metodei celor mai mici pătrate rezultă următorul sistem de ecuaţii normale:

- cînd perechile de valori corelate x_i, y_i sînt neponderate:

$$\begin{aligned} n\,a + b\sum x_i + c\sum x_i^2 &= \sum \log y_i \\ a\sum x_i + b\sum x_i^2 + c\sum x_i^3 &= \sum x_i \log y_i \\ a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 &= \sum x_i^2 \log y_i \end{aligned} \qquad (15.33)$$

- cînd perechile de valori corelate x_i, y_i sînt ponderate cu frecvenţa n_i :

$$\begin{aligned} a\sum n_i + b\sum x_i n_i + c\sum x_i^2 n_i &= \sum (\log y_i) n_i \\ a\sum x_i n_i + b\sum x_i^2 n_i + c\sum x_i^3 n_i &= \sum x_i (\log y_i) n_i \\ a\sum x_i^2 n_i + b\sum x_i^3 n_i + c\sum x_i^4 n_i &= \sum x_i^2 (\log y_i) n_i \end{aligned} \qquad (15.34)$$

- cînd valorile celor două variabile corelate X şi Y se prezintă sub forma unui tabel de corelaţie:

$$\begin{aligned} a\sum\sum n_{ij} + b\sum x_i n_{i.} + c\sum x_i^2 n_{i.} &= \sum (\log y_j) n_{.j} \\ a\sum x_i n_{i.} + b\sum x_i^2 n_{i.} + c\sum x_i^3 n_{i.} &= \sum\sum x_i (\log y_j) n_{ij} \\ a\sum x_i^2 n_{i.} + b\sum x_i^3 n_{i.} + c\sum x_i^4 n_{i.} &= \sum\sum x_i^2 (\log y_j) n_{ij} \end{aligned} \qquad (15.35)$$

Formulele de calcul simplificat ale raportului de corelaţie vor fi:

- în cazul perechilor de valori x_i, y_i neponderate:

$$\eta = \sqrt{\frac{a\sum \log y_i + b\sum x_i \log y_i + c\sum x_i^2 \log y_i - \frac{(\sum \log y_i)^2}{n}}{\sum (\log y_i)^2 - \frac{(\sum \log y_i)^2}{n}}} \qquad (15.36)$$

- în cazul perechilor de valori x_i, y_i ponderate cu n_i:

$$\eta = \sqrt{\frac{a\sum n_i \log y_i + b\sum n_i x_i \log y_i + c\sum n_i x_i^2 \log y_i - \frac{(\sum n_i \log y_i)^2}{\sum n_i}}{\sum n_i (\log y_i)^2 - \frac{(\sum n_i \log y_i)^2}{\sum n_i}}} \qquad (15.37)$$

- în cazul prezentării celor două variabile sub forma unui tabel de corelaţie:

$$\eta = \sqrt{\frac{a\sum n_{.j} \log y_j + b\sum\sum n_{ij} x_i \log y_j + c\sum\sum n_{ij} x_i^2 \log y_j - \frac{(\sum n_{.i} \log y_i)^2}{\sum\sum n_{ij}}}{\sum n_{.j} (\log y_j)^2 - \frac{(\sum n_{.i} \log y_i)^2}{\sum\sum n_{ij}}}} \qquad (15.38)$$

O altă formă a parabolei de gradul doi semilogaritmică se obţine prin substituirea valorilor lui x_i cu logaritmii acestora din formula (15.1) :

$$y = a + b \log x_i + c (\log x_i)^2$$

In cazul perechilor de valori x_i, y_i neponderate, parametrii parabolei semilogaritmice se determină prin rezolvarea sistemului de ecuaţii normale (15.39):

$$\begin{aligned} \sum y_i &= n\,a + b \sum \log x_i + c \sum (\log x_i)^2 \\ \sum y_i \log x_i &= a \sum \log x_i + b \sum (\log x_i)^2 + c \sum (\log x_i)^3 \\ \sum y_i (\log x_i)^2 &= a \sum (\log x_i)^2 + b \sum (\log x_i)^3 + c \sum (\log x_i)^4 \end{aligned} \qquad (15.39)$$

- în cazul prezentării perechilor de valori x_i, y_i ponderate cu n_i :

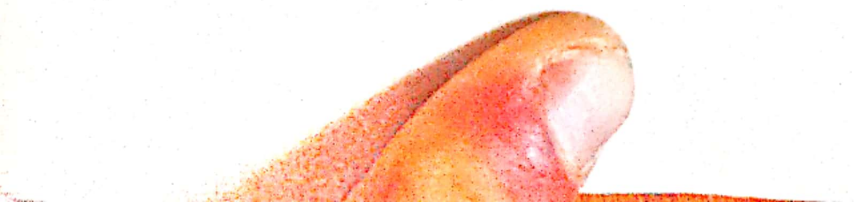
$$\begin{aligned} \sum y_i n_i &= a \sum n_i + b \sum n_i \log x_i + c \sum n_i (\log x_i)^2 \\ \sum n_i y_i \log x_i &= a \sum n_i \log x_i + b \sum n_i (\log x_i)^2 + c \sum n_i (\log x_i)^3 \\ \sum n_i y_i (\log x_i)^2 &= a \sum n_i (\log x_i)^2 + b \sum n_i (\log x_i)^3 + c \sum n_i (\log x_i)^4 \end{aligned} \qquad (15.40)$$

- în cazul prezentării celor două variabile sub forma unui tabel de corelaţie:

$$\begin{aligned} \sum y_j n_{\cdot j} &= a \sum\sum n_{ij} + b \sum n_{i\cdot} \log x_i + c \sum n_{i\cdot} (\log x_i)^2 \\ \sum\sum n_{ij} y_j \log x_i &= a \sum n_{i\cdot} \log x_i + b \sum n_{i\cdot} (\log x_i)^2 + c \sum n_{i\cdot} (\log x_i)^3 \\ \sum\sum n_{ij} y_j (\log x_i)^2 &= a \sum n_{i\cdot} (\log x_i)^2 + b \sum n_{i\cdot} (\log x_i)^3 + c \sum n_{i\cdot} (\log x_i)^4 \end{aligned} \qquad (15.41)$$

Formulele de calcul simplificat ale raportului de corelaţie au următoarele expresii în cele trei cazuri posibile de prezentare a datelor iniţiale :

- în cazul prezentării perechilor de valori corelate



x_i, y_i neponderate:

$$\eta = \sqrt{\frac{a\sum y_i + b\sum y_i \log x_i + c\sum y_i (\log x_i)^2 - \frac{(\sum y_i)^2}{n}}{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}} \qquad (15.42)$$

- în cazul prezentării perechilor de valori x_i, y_i ponderate cu n_i:

$$\eta = \sqrt{\frac{a\sum y_i n_i + b\sum n_i y_i \log x_i + c\sum n_i y_i (\log x_i)^2 - \frac{(\sum y_i n_i)^2}{\sum n_i}}{\sum y_i^2 n_i - \frac{(\sum y_i n_i)^2}{\sum n_i}}} \qquad (15.43)$$

- în cazul prezentării celor două variabile sub forma unui tabel de corelaţie :

$$\eta = \sqrt{\frac{a\sum y_j n_{\cdot j} + b\sum\sum n_{ij} y_j \log x_i + c\sum\sum n_{ij} y_j (\log x_i)^2 - \frac{\sum (y_i n_{\cdot i})^2}{\sum\sum n_{ij}}}{\sum y_j^2 n_{\cdot j} - \frac{(\sum y_i n_{\cdot i})^2}{\sum\sum n_{ij}}}} \qquad (15.44)$$

Dacă în formula (15.1) se substituie ambele variabile cu logaritmii lor se obţine parabola de gradul doi dublulogaritmică :

$$\log y_i = a + b(\log x_i) + c\,(\log x_i)^2$$

Expresia geometrică a acestei funcţii se prezintă în fig. 15.4.

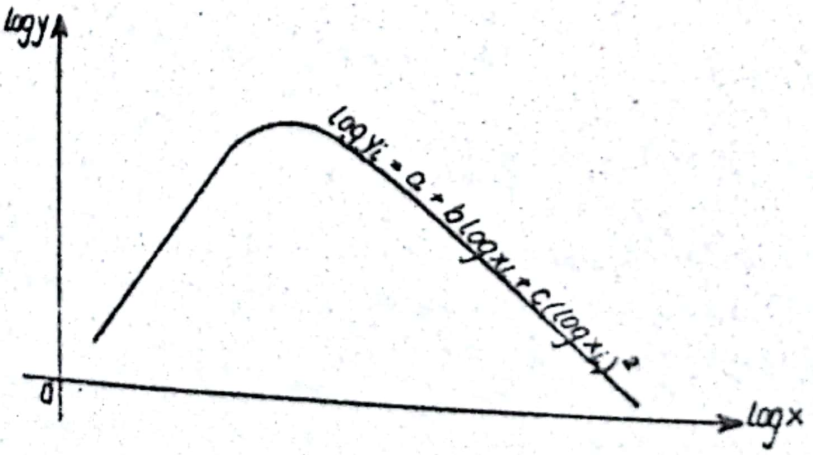


Fig 15.4. PARABOLA DE GRADUL DOI DUBLULOGARITMICĂ

Se observă în fig. 15.4 că parabola dublulogaritmică este o funcţie asimetrică şi prezintă un punct extrem.

Pentru parabola dublulogaritmică de gradul doi (15.7) sistemul de ecuaţii normale capătă următoarele expresii:

- în cazul prezentării perechilor de valori corelate x_i, y_i neponderate:

$$n\,a + b\sum \log x_i + c\sum (\log x_i)^2 = \sum \log y_i \qquad (15.45)$$

$$a\sum \log x_i + b\sum (\log x_i)^2 + c\sum (\log x_i)^3 = \sum (\log x_i)(\log y_i)$$

$$a\sum (\log x_i)^2 + b\sum (\log x_i)^3 + c\sum (\log x_i)^4 = \sum (\log x_i)^2 (\log y_i)$$

- dacă perechile de valori corelate x_i, y_i sînt ponderate cu n_i:

$$a\sum n_i + b\sum n_i \log x_i + c\sum n_i (\log x_i)^2 = \sum n_i \log y_i \qquad (15.46)$$

$$a\sum n_i \log x_i + b\sum n_i (\log x_i)^2 + c\sum n_i (\log x_i)^3 = \sum n_i (\log x_i)(\log y_i)$$

$$a\sum n_i (\log x_i)^2 + b\sum n_i (\log x_i)^3 + c\sum n_i (\log x_i)^4 = \sum n_i (\log x_i)^2 (\log y_i)$$

- dacă variabilele X şi Y sînt prezentate sub forma unui tabel de corelaţie:

$$a\sum\sum n_{ij} + b\sum n_{i.} \log x_i + c\sum n_{i.} (\log x_i)^2 = \sum n_{.j} \log y_j \qquad (15.47)$$

$$a\sum n_{i.} \log x_i + b\sum n_{i.} (\log x_i)^2 + c\sum n_{i.} (\log x_i)^3 = \sum\sum n_{ij} (\log x_i)(\log y_j)$$

$$a\sum n_{i.} (\log x_i)^2 + b\sum n_{i.} (\log x_i)^3 + c\sum n_{i.} (\log x_i)^4 = \sum\sum n_{ij} (\log x_i)(\log y_j)$$

Formulele de calcul simplificat ale raportului de corelaţie simplă parabolică se prezintă astfel:

- în cazul perechilor de valori corelate x_i, y_i neponderate:

$$\eta = \sqrt{\frac{a\sum \log y_i + b\sum (\log x_i)(\log y_i) + c\sum (\log x_i)^2 (\log y_i) - \frac{(\sum \log y_i)^2}{n}}{\sum (\log y_i)^2 - \frac{(\sum \log y_i)^2}{n}}} \qquad (15.48)$$

- în cazul în care perechile de valori x_i, y_i sînt ponderate cu frecvenţa n_i:

$$\eta = \sqrt{\frac{a\sum n_i \log y_i + b\sum n_i (\log x_i)(\log y_i) + c\sum n_i (\log x_i)^2 (\log y_i) - \frac{(\sum \log y_i)^2}{\sum n_i}}{\sum (\log y_i)^2 n_i - \frac{(\sum n_i \log y_i)^2}{\sum n_i}}} \qquad (15.49)$$

- în cazul în care variabilele X şi Y se prezintă sub forma unui tabel de corelaţie:

$$\eta = \sqrt{\frac{a\sum n_{\cdot j} \log y_j + b\sum\sum (\log x_i)(\log y_j) n_{ij} + c\sum\sum (\log x_i)^2 (\log y_j) n_{ij} - \frac{(\sum n_{\cdot j} \log y_j)^2}{\sum\sum n_{ij}}}{\sum (\log y_j)^2 n_{\cdot j} - \frac{(\sum n_{\cdot j} \log y_j)^2}{\sum\sum n_{ij}}}} \qquad (15.50)$$

Legătura simplă dintre fenomene poate îmbrăca şi forma unei parabole de gradul trei.

Parabola de gradul trei $y = a+b\,x+c\,x^2+d\,x^3$ este o curbă asimetrică şi prezintă două puncte de extrem: unul de maxim şi unul de minim (fig. 15.5) şi un singur punct de inflexiune. Parabola de gradul trei depinde deci de patru parametri a,b,c,d şi va fi necesar să se selecţioneze patru puncte echidistante pentru estimarea celor patru parametri ai curbei.

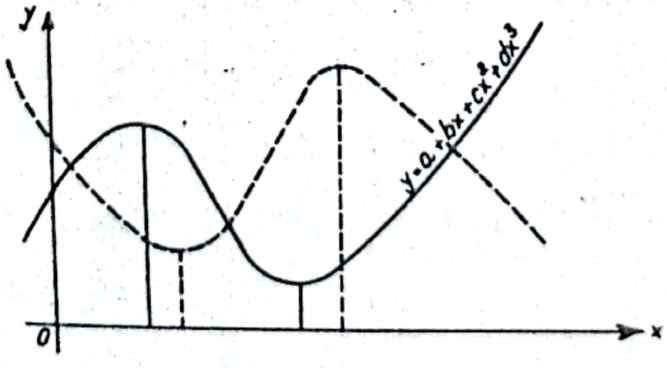


Fig. 15.5 PARABOLA DE GRADUL TREI

Cînd $d>0$ atunci apare întîi punctul de maxim şi apoi cel de minim (15.5 - curba A), iar cînd $d<0$ punctul de minim precede pe cel de maxim (fig. 15.5 - curba întreruptă).

Parametrii parabolei de gradul trei se obţin tot cu ajutorul metodei celor mai mici pătrate:

$$S = \sum(y_i - a - b\,x_i - c\,x_i^2 - d\,x_i^3)^2 = \text{minim}$$

$$\frac{\partial S}{\partial a} = 2\sum(y_i - a - b\,x_i - c\,x_i^2 - d\,x_i^3)(-1) = 0$$

$$\frac{\partial S}{\partial b} = 2\sum(y_i - a - b\,x_i - c\,x_i^2 - d\,x_i^3)(-x_i) = 0$$

$$\frac{\partial S}{\partial c} = 2\sum(y_i - a - b\,x_i - c\,x_i^2 - d\,x_i^3)(-x_i^2) = 0$$

$$\frac{\partial S}{\partial d} = 2\sum(y_i - a - b\,x_i - c\,x_i^2 - d\,x_i^3)(-x_i^3) = 0$$

Prin transformări elementare sistemul devine:

$$\begin{aligned}
&n\,a + b\sum x_i - c\sum x_i^2 + d\sum x_i^3 = \sum y_i \\
&a\sum x_i + b\sum x_i^2 + c\sum x_i^3 + d\sum x_i^4 = \sum x_i y_i \\
&a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 + d\sum x_i^5 = \sum x_i^2 y_i \\
&a\sum x_i^3 + b\sum x_i^4 + c\sum x_i^5 + d\sum x_i^6 = \sum x_i^3 y_i
\end{aligned} \qquad (15.51)$$

In cazul în care frecvenţele n_i ale perechilor de valori corelate x_i, y_i sînt diferite de unitate, sistemul (15.51) devine:

$$\begin{aligned}
&a\sum n_i + b\sum x_i n_i + c\sum x_i^2 n_i + d\sum x_i^3 n_i = \sum y_i n_i \\
&a\sum x_i n_i + b\sum x_i^2 n_i + c\sum x_i^3 n_i + d\sum x_i^4 n_i = \sum x_i y_i n_i \\
&a\sum x_i^2 n_i + b\sum x_i^3 n_i + c\sum x_i^4 n_i + d\sum x_i^5 n_i = \sum x_i^2 y_i n_i \\
&a\sum x_i^3 n_i + b\sum x_i^4 n_i + c\sum x_i^5 n_i + d\sum x_i^6 n_i = \sum x_i^3 y_i n_i
\end{aligned} \qquad (15.52)$$

Dacă variantele variabilelor corelate se prezintă sub forma unui tabel de corelaţie, atunci sistemul de ecuaţii normale ia forma:

$$
\begin{aligned}
a\sum\sum n_{ij} + b\sum x_i n_{i\cdot} + c\sum x_i^2 n_{i\cdot} + d\sum x_i^3 n_{i\cdot} &= \sum y_j n_{\cdot j} \\
a\sum x_i n_{i\cdot} + b\sum x^2 n_{i\cdot} + c\sum x_i^3 n_{i\cdot} + d\sum x_i^4 n_{i\cdot} &= \sum\sum x_i y_j n_{ij} \\
a\sum x_i^2 n_{i\cdot} + b\sum x_i^3 n_{i\cdot} + c\sum x_i^4 n_{i\cdot} + d\sum x_i^5 n_{i\cdot} &= \sum\sum x_i y_j n_{ij} \\
a\sum x_i^3 n_{i\cdot} + b\sum x_i^4 n_{i\cdot} + c\sum x_i^5 n_{i\cdot} + d\sum x_i^6 n_{i\cdot} &= \sum\sum x_i y_j n_{ij}
\end{aligned}
\qquad (15.53)
$$

Prin rezolvarea acestor sisteme, cu ajutorul unei dintre metodele cunoscute, se obţin valorile celor patru parametri ai ecuaţiei de regresie în fiecare dintre cele trei modalităţi de prezentare a datelor statistice.

Intensitatea legăturii dintre variabile se măsoară cu ajutorul raportului de corelaţie, formula generală (15.24).

Cunoscînd parametrii parabolei de gradul trei se poate obţine o formulă de calcul simplicat a raportului de corelaţie simplă după modelul parabolei de gradul trei:

$$
\eta = \sqrt{\frac{a\sum y_i + b\sum x_i y_i + c\sum x_i^2 y_i + d\sum x_i^3 y_i - \frac{(\sum y_i)^2}{n}}{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}} \qquad (15.54)
$$

In cazul ponderării perechilor de valori ale variabilelor x_i, y_i cu frecvenţa n_i, raportul de corelaţie are următoarea expresie:

$$
\eta = \sqrt{\frac{a\sum y_i n_i + b\sum x_i y_i n_i + c\sum x_i^2 y_i n_i + d\sum x_i^3 y_i n_i - \frac{(\sum y_i n_i)^2}{\sum n_i}}{\sum y_i^2 n_i - \frac{(\sum y_i n_i)^2}{\sum n_i}}} \qquad (15.55)
$$

Dacă cele două variabile corelate X şi Y se prezintă sub forma unui tabel de corelaţie η devine:

$$
\eta = \sqrt{\frac{a\sum y_j n_{\cdot j} + b\sum\sum x_i y_j n_{ij} + c\sum\sum x_i^2 y_j n_{ij} + d\sum\sum x_i y_j n_{ij} - \frac{(\sum y_i n_{\cdot i})^2}{\sum\sum n_{ij}}}{\sum y_j^2 n_{\cdot j} - \frac{(\sum y_i n_{\cdot i})^2}{\sum\sum n_{ij}}}} \qquad (15.56)
$$

Dacă în locul valorilor lui y din cadrul ecuației parabolei de gradul trei se utilizează logaritmii valorilor respective, adică $\log y_i$, atunci parabola de gradul trei semilogaritmică va fi descrisă de următoarea ecuație:

$$\log y_i = a + b x_i + c x_i^2 + d x_i^3$$

Prin aplicarea metodei celor mai mică pătrate se obțin următoarele trei forme ale sistemului de ecuații normale corespunzător celor trei modalități de sistematizare a datelor inițiale:

- când perechile de valori x_i, y_i sînt neponderate:

$$\begin{aligned}
&n a + b\sum x_i + c\sum x_i^2 + d\sum x_i^3 = \sum \log y_i \\
&a\sum x_i + b\sum x_i^2 + c\sum x_i^3 + d\sum x_i^4 = \sum x_i \log y_i \\
&a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 + d\sum x_i^5 = \sum x_i^2 \log y_i \\
&a\sum x_i^3 + b\sum x_i^4 + c\sum x_i^5 + d\sum x_i^6 = \sum x_i^3 \log y_i
\end{aligned} \qquad (15.57)$$

- când perechile de valori x_i, y_i sînt ponderate cu frecvența n_i:

$$\begin{aligned}
&a\sum n_i + b\sum x_i n_i + c\sum x_i^2 n_i + d\sum x_i^3 n_i + d\sum x_i^4 n_i = \sum n_i \log y_i \\
&a\sum x_i n_i + b\sum x_i^2 n_i + c\sum x_i^3 n_i + d\sum x_i^4 n_i = \sum n_i x_i \log y_i \\
&a\sum x_i^2 n_i + b\sum x_i^3 n_i + c\sum x_i^4 n_i + d\sum x_i^5 n_i = \sum n_i x_i^2 \log y_i \\
&a\sum x_i^3 n_i + b\sum x_i^4 n_i + c\sum x_i^5 n_i + d\sum x_i^6 n_i = \sum n_i x_i^3 \log y_i
\end{aligned} \qquad (15.58)$$

- când valorile celor două variabile se prezintă sub forma unui tabel de corelație:

$$\begin{aligned}
&a\sum\sum n_{ij} + b\sum x_i n_{i.} + c\sum x_i^2 n_{i.} + d\sum x_i^3 n_{i.} = \sum n_{.j} \log y_j \\
&a\sum x_i n_{i.} + b\sum x_i^2 n_{i.} + c\sum x_i^3 n_{i.} + d\sum x_i^4 n_{i.} = \sum\sum n_{ij} x_i \log y_j
\end{aligned} \qquad (15.59)$$

$$a\sum x_i^2 n_{i\cdot} + b\sum x_i^3 n_{i\cdot} + c\sum x_i^4 n_{i\cdot} + d\sum x_i^5 n_{i\cdot} = \sum\sum n_{ij} x_i^2 \log y_j$$

$$a\sum x_i^3 n_{i\cdot} + b\sum x_i^4 n_{i\cdot} + c\sum x_i^5 n_{i\cdot} + d\sum x_i^6 n_{i\cdot} = \sum\sum n_{ij} x_i^3 \log y_j$$

In primul caz, formula raportului de corelaţie se obţine pe baza formulei:

$$\eta = \sqrt{\frac{a\sum \log y_i + b\sum x_i \log y_i + c\sum x_i^2 \log y_i + d\sum x_i^3 \log y_i - \frac{(\sum \log y_i)^2}{n}}{\sum (\log y_i)^2 - \frac{(\sum \log y_i)^2}{n}}} \qquad (15.60)$$

Dacă perechile de valori corelate x_i, y_i sînt ponderate cu frecvenţa n_i:

$$\eta = \sqrt{\frac{a\sum n_i \log y_i + b\sum n_i x_i \log y_i + c\sum n_i x_i^2 \log y_i + d\sum n_i x_i^3 \log y_i - \frac{(\sum n_i \log y_i)^2}{\sum n_i}}{\sum (\log y_i)^2 n_i - \frac{(\sum n_i \log y_i)^2}{\sum n_i}}} \qquad (15.61)$$

Dacă valorile celor două variabile se prezintă sub forma unui tabel de corelaţie:

$$\eta = \sqrt{\frac{a\sum n_{\cdot j} \log y_j + b\sum\sum n_{ij} x_i \log y_j + c\sum\sum n_{ij} x_i^2 \log y_j + d\sum\sum n_{ij} x_i^3 \log y_j - \frac{(\sum n_j \log y_j)^2}{\sum\sum n_{ij}}}{\sum n_{\cdot j} (\log y_j)^2 - \frac{(\sum n_{\cdot j} \log y_j)^2}{\sum\sum n_{ij}}}} \qquad (15.62)$$

Dacă se substituie în ecuaţia parabolei de gradul trei valorile x_i cu $\log x_i$ atunci ecuaţia parabolei devine:

$$y = a + b \log x_i + c (\log x_i)^2 + d(\log x_i)^3$$

Aplicînd metoda celor mai mici pătrate pentru determinarea parametrilor de regresie a, b, c şi d, se obţine sistemul de ecuaţii normale:

$$n\,a+b\sum\log x_i + c\sum(\log x_i)^2+d\sum(\log x_i)^3 = \sum y_i \qquad (15.63)$$

$$a\sum\log x_i+b\sum(\log x_i)^2+c\sum(\log x_i)^3+d\sum(\log x_i)^4=\sum y_i\log x_i$$

$$a\sum(\log x_i)^2+b\sum(\log x_i)^3+c\sum(\log x_i)^4+d\sum(\log x_i)^5=\sum y_i(\log x_i)^2$$

$$a\sum(\log x_i)^3+b\sum(\log x_i)^4+c\sum(\log x_i)^5+d\sum(\log x_i)^6=\sum y_i(\log x_i)^3$$

Dacă perechile de valori $x_i y_i$ sînt ponderate cu n_i sistemul de ecuaţii normale ia forma:

$$a\sum n_i+b\sum n_i\log x_i+c\sum n_i(\log x_i)^2+d\sum n_i(\log x_i)^3=\sum y_i n_i \qquad (15.64)$$

$$a\sum n_i\log x_i+b\sum n_i(\log x_i)^2+c\sum n_i(\log x_i)^3+d\sum n_i(\log x_i)^4=\sum n_i y_i\log x_i$$

$$a\sum n_i(\log x_i)^2+b\sum(\log x_i)^3+c\sum n_i(\log x_i)^4+d\sum n_i(\log x_i)^5=\sum n_i y_i(\log x_i)^2$$

$$a\sum n_i(\log x_i)^3+b\sum n_i(\log x_i)^4+c\sum n_i(\log x_i)^5+d\sum n_i(\log x_i)^6=\sum n_i y_i(\log x_i)^3$$

In cazul unui tabel de corelaţie sistemul de ecuaţii normale se scrie astfel:

$$a\sum\sum n_{ij}+b\sum n_{i.}\log x_i+c\sum n_{i.}(\log x_i)^2+d\sum n_{i.}(\log x_i)^3=\sum y_j n_{.j} \qquad (16.65)$$

$$a\sum n_{i.}\log x_i+b\sum n_{i.}(\log x_i)^2+c\sum n_{i.}(\log x_i)^3+d\sum n_{i.}(\log x_i)^4=$$

$$=\sum\sum n_{ij}y_j\log x_i$$

$$a\sum n_{i.}(\log x_i)^2+b\sum n_{i.}(\log x_i)^3+c\sum n_{i.}(\log x_i)^4+d\sum n_{i.}(\log x_i)^5=$$

$$=\sum\sum n_{ij}y_j(\log x_i)^2$$

$$a\sum n_{i.}(\log x_i)^3+b\sum n_{i.}(\log x_i)^4+c\sum n_{i.}(\log x_i)^5+d\sum n_{i.}(\log x_i)^6=$$

$$=\sum\sum n_{ij}y_j(\log x_i)^3$$

In funcţie de modul de prezentare a valorilor celor două variabile raportul de corelaţie în cazul regresiei după o parabolă de gradul trei semilogaritmică va avea următoarele expresii:

- cînd perechile de valori corelate x_i, y_i se întîlnesc o singură dată:

$$\eta = \sqrt{\frac{a\sum y_i + b\sum y_i \log x_i + c\sum y_i (\log x_i)^2 + d\sum y_i (\log x_i)^3 - \frac{(\sum y_i)^2}{n}}{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}} \qquad (15.66)$$

- cînd perechile de valori x_i, y_i sînt ponderate cu n_i:

$$\eta = \sqrt{\frac{a\sum y_i n_i + b\sum n_i y_i \log x_i + c\sum n_i y_i (\log x_i)^2 + d\sum n_i y_i (\log x_i)^3 - \frac{(\sum y_i n_i)^2}{\sum n_i}}{\sum y_i^2 n_i - \frac{(\sum y_i n_i)^2}{\sum n_i}}} \qquad (15.67)$$

- cînd valorile celor două variabile se prezintă sub forma unui tabel de corelaţie:

$$\eta = \sqrt{\frac{a\sum y_j n_{\cdot j} + b\sum\sum n_{ij} y_j \log x_i + c\sum\sum n_{ij} y_i (\log x_i)^2 + d\sum\sum n_{ij} (\log x_i)^3 - \frac{(\sum y_i n_{ii})^2}{\sum\sum n_{ii}}}{\sum y_j^2 n_{\cdot j} - \frac{(\sum y_i n_{\cdot i})^2}{\sum\sum n_{ij}}}} \qquad (15.68)$$

In cazul parabolei de gradul trei logaritmică:

$$\log y = a + b \log x + c (\log x)^2 + a (\log x)^3$$

In cele trei cazuri posibile de prezentare a datelor iniţiale pentru determinarea parametrilor ecuaţiei de regresie se vor rezolva următoarele trei sisteme de ecuaţii normale:

$$n\,a + b\sum \log x_i + c\sum (\log x_i)^2 + d\sum (\log x_i)^3 = \sum \log y_i \qquad (15.69)$$

$$a\sum \log x_i + b\sum (\log x_i)^2 + c\sum (\log x_i)^3 + d\sum (\log x_i)^4 = \sum (\log x_i)(\log y_i)$$

$$a\sum (\log x_i)^2 + b\sum (\log x_i)^3 + c\sum (\log x_i)^4 + d\sum (\log x_i)^5 = \sum (\log x_i)^2 (\log y_i)$$

$$a\sum (\log x_i)^3 + b\sum (\log x_i)^4 + c\sum (\log x_i)^5 + d\sum (\log x_i)^6 = \sum (\log x_i)^3 (\log y_i)$$

şi

$$a\sum n_i + b\sum n_i \log x_i + c\sum n_i (\log x_i)^2 + d\sum n_i (\log x_i)^3 = \sum n_i \log y_i \qquad (15.70)$$

$$a\sum n_i \log x_i + b\sum n_i (\log x_i)^2 + c\sum n_i (\log x_i)^3 + d\sum n_i (\log x_i)^4 =$$

$$= \sum n_i (\log x_i)(\log y_i)$$

$$a\sum n_i (\log x_i)^2 + b\sum n_i (\log x_i)^3 + c\sum n_i (\log x_i)^4 + d\sum n_i (\log x_i)^5 =$$

$$= \sum n_i (\log x_i)^2 (\log y_i)$$

$$a\sum n_i(\log x_i)^3+b\sum n_i(\log x_i)^4+c\sum n_i(\log x_i)^5+d\sum n_i(\log x_i)^6 =$$
$$= \sum n_i(\log x_i)^3(\log y_i)$$

şi

$$a\sum n_{ij}+b\sum n_{i\cdot}\log x_i+c\sum n_{i\cdot}(\log x_i)^2+d\sum n_{i\cdot}(\log x_i)^3 = \sum n_{\cdot j}\log y_j \qquad (15.71)$$

$$a\sum n_{i\cdot}\log x_i+b\sum n_{i\cdot}(\log x_i)^2+c\sum n_{i\cdot}(\log x_i)^3+d\sum n_{i\cdot}(\log x_i)^4 =$$
$$=\sum\sum n_{ij}(\log x_i)(\log y_i)$$

$$a\sum n_{i\cdot}(\log x_i)^2+b\sum n_{i\cdot}(\log x_i)^3+c\sum n_{i\cdot}(\log x_i)^4+d\sum n_{i\cdot}(\log x_i)^5=$$
$$=\sum\sum n_{ij}(\log x_i)^2(\log y_j)$$

$$a\sum n_{i\cdot}(\log x_i)^3+b\sum n_{i\cdot}(\log x_i)^4+c\sum n_{i\cdot}(\log x_i)^5+d\sum n_{i\cdot}(\log x_i)^6=$$
$$=\sum\sum n_{ij}(\log x_i)^3(\log y_j)$$

Raportul de corelaţie corespunzător celor trei modalităţi de prezentare a datelor se determină astfel:

- cînd perechile de valori corelate x_i, y_i se întîlnesc o singură dată:

$$\eta=\sqrt{\frac{a\sum\log y_i+b\sum(\log x_i)(\log y_i)+c\sum(\log x_i)^2(\log y_i)+d\sum(\log x_i)^3(\log y_i)-\frac{\sum\log y_i)^2}{n}}{\sum(\log y_i)^2-\frac{(\sum\log y_i)^2}{n}}} \qquad (15.72)$$

- cînd perechile de valori x_i, y_i sînt ponderate cu frecvenţa n_i:

$$\eta=\sqrt{\frac{a\sum n_i\log y_i+b\sum n_i(\log x_i)(\log y_i)+c\sum n_i(\log x_i)^3(\log y_i)+d\sum n_i(\log x_i)(\log y_i)-\frac{(\sum\log y_i)^2}{\sum n_i}}{\sum n_i(\log y_i)^2-\frac{(\sum n_i\log y_i)^2}{\sum n_i}}} \qquad (15.73)$$

- cînd variabilele X şi Y se prezintă sub forma unui tabel de corelaţie:

$$\eta = \sqrt{\frac{a\sum n_{\cdot j}\log y_j + b\sum\sum n_{ij}(\log x_i)(\log y_j) + c\sum\sum n_{ij}(\log x_i)^2(\log y_j) + d\sum\sum n_{ij}(\log x_i)^3(\log y_j) - \frac{(\sum n_{\cdot j}\log y_j)^2}{\sum\sum n_{ij}}}{\sum n_{\cdot j}(\log y_j)^2 - \frac{(\sum n_{\cdot j}\log y_j)^2}{\sum\sum n_{ij}}}} \qquad (15.74)$$

Desigur că pentru parabolele de grad superior indicatorii de regresie şi corelaţie simplă se vor calcula şi analiza în mod asemănător. Cu cît creşte gradul parabolei cu atît creşte şi numărul punctelor de extrem. Astfel, în cazul unei drepte (care poate fi considerată drept o parabolă de gradul întîi) depinde de doi parametri şi nu are nici un punct de extrem. Parabola de gradul depinde de doi parametri şi are un punct de extrem. Parabola de gradul trei depinde de patru parametri şi are două puncte de extrem. In genaral, parabola de gradul n depinde de n+1 parametri şi are n-1 puncte de extrem şi n-2 puncte de inflexiune. In practica statistică se tinde a se folosi curbe de regresie cît mai simple şi de aceea numai în cazuri rare se recurge la parabole de un grad mai mare decît trei.

15.1.2. Legături de tip hiperbolic

Legăturile simple dintre fenomenele reale pot să ia forma unei curbe hiperbolice. Ecuaţia hiperbolei are mai multe expresii în funcţie de caracteristicile ei principale.

Astfel, dependenţa de tip hiperbolic poate fi descrisă de ecuaţia următoare:

$$y = a + \frac{b}{x} \qquad (15.75)$$

sau

$$y = \frac{a\,x + b}{x}$$

unde $(0 < x < \infty)$.

Graficul funcţiei hiperbolice (15.9) se prezintă în fig. 15.6.

Curba descrisă de ecuaţia (15.75) are două asimptote: (vezi fig. 15.6) una din orizontală $y = a$ şi una din verticală

$x = 0$. Pentru $b>0$, curba este descrescătoare şi tinde asimptotic către a , iar dacă $b<0$ atunci hiperbola este descrisă de o curbă crescătoare care tinde asimptotic spre o valoare limită superioară a cînd x tinde către $+\infty$. Cînd $b<0$ legătura dintre variabile este directă sau pozitivă, iar cînd $b>0$ o legătură dintre variabile este inversă sau negativă.

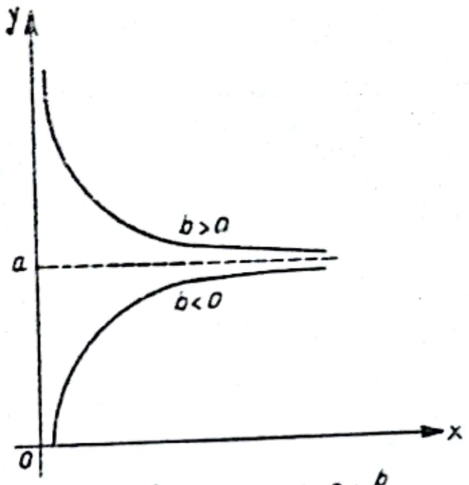


Fig.15.6. HIPERBOLA $y = a + \frac{b}{x}$

Dependenţa de tip hiperbolic poate lua şi o altă formă şi anume:

$$y = \frac{1}{a + b\,x} \quad \text{unde} \quad \left(-\frac{a}{b} < x < \infty\right) \qquad (15.76)$$

Graficul acestei funcţii se prezintă în fig. 15.7.

După cum se observă, din fig.15.7, cele două curbe hiperbolice au tot două asimptote: una orizontală $y = 0$ şi alta verticală $x = -\frac{a}{b}$. In aplicaţiile practice se întîlneşte doar cazul (A) şi numai partea din curbă situată în cadranul întîi, corespunzătoare unor valori pozitive ale celor două variabile. Curba din fig. 15.7(A) este o curbă descrescătoare ce corespunde unor legături inverse între variabilele corelate.

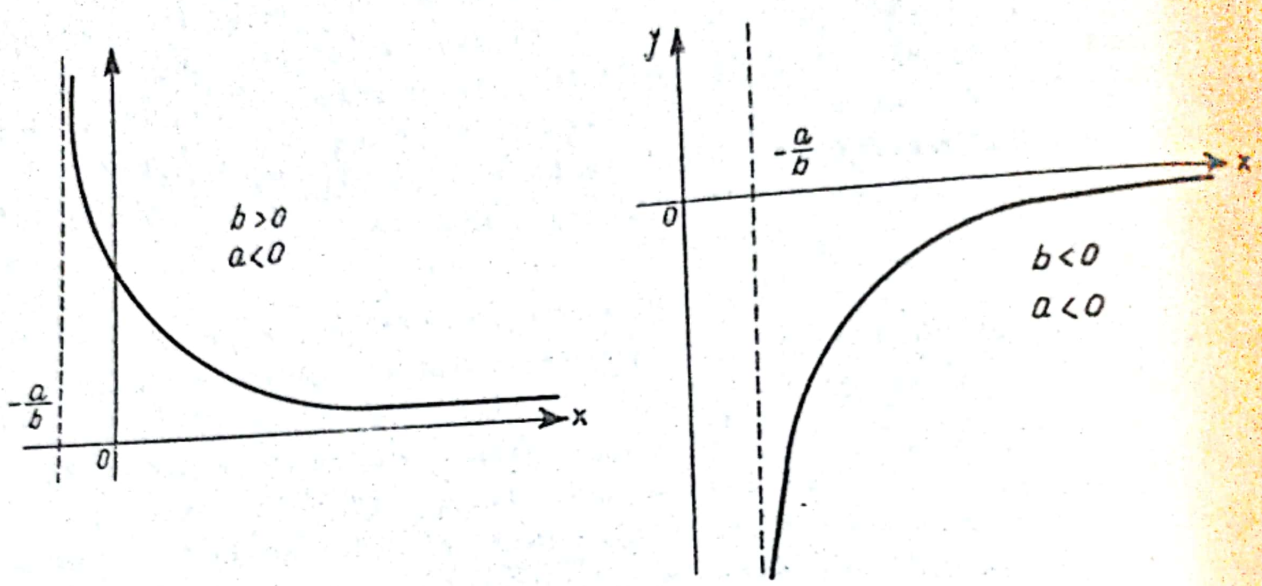


Fig. 15.7. HIPERBOLA $y = \frac{1}{a + bx}$

Dependenţa de tip hiperbolic poate fi descrisă şi de ecuaţia:

$$y = \frac{1}{a + \frac{b}{x}} \qquad (15.77)$$

sau

$$y = \frac{x}{a\,x + b} \quad \text{unde} \quad \left(-\frac{b}{a} < x < \infty\right)$$

Curba ce descrie dependenţa hiperbolică (15.77) se prezintă în fig. 15.8.

Hiperbola din fig. 15.8 are două asimptote: una orizontală $y = \frac{1}{a}$ şi una verticală $x = -\frac{b}{a}$. Curba din fig. 15.8 (A) prezintă o legătură inversă, iar cea din fig. 15.8 (B) o legătură directă între variabilele corelate.

Parametrii ecuaţiei de regresie de tip hiperbolic (15.75) se determină cu ajutorul metodei celor mai mică pătrate:

$$y_{x_i} = a + \frac{b}{x_i}$$

Suma pătratelor minime ale abaterilor valorilor empirice y_i faţă de valorile teoretice y_{x_i} este:

$$S = \sum (y_i - y_{x_i})^2 = \text{minim}$$

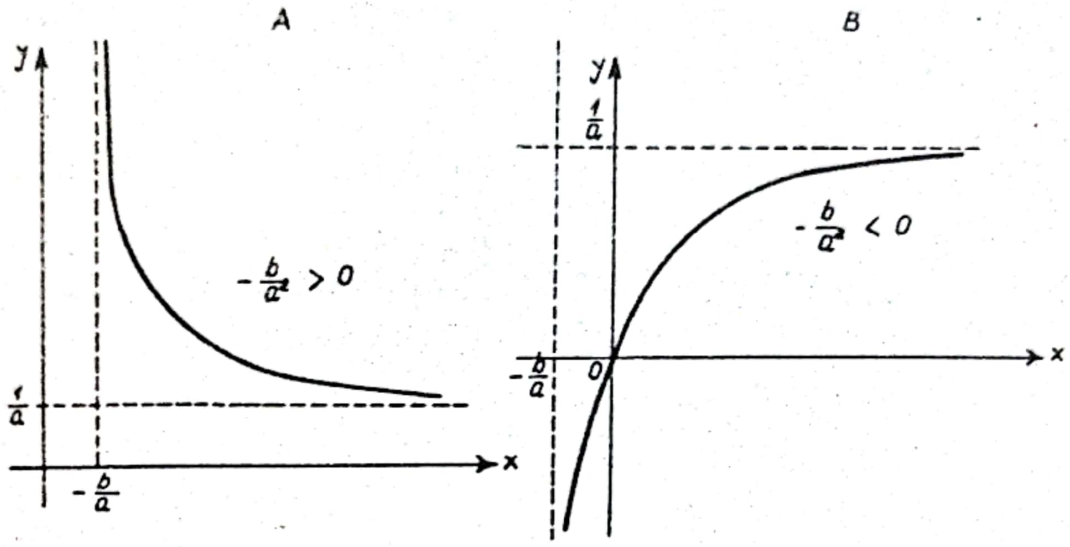


Fig.15.8. HIPERBOLA $y = \frac{1}{a + \frac{b}{x}}$

Substituind în expresia de mai sus pe y_x cu expresia lui din (15.75) rezultă:

$$S = \sum (y_i - a - \frac{b}{x_i})^2 = \text{minim} \qquad (15.78)$$

Derivînd funcţia S în raport cu a şi b şi egalînd derivatele parţiale cu zero, se obţine sistemul de ecuaţii normale:

$$\frac{\partial S}{\partial a} = 2 \sum (y_i - a - \frac{b}{x_i}) \cdot (-1) = 0$$

$$\frac{\partial S}{\partial b} = 2 \sum (y_i - a - \frac{b}{x_i})(-\frac{1}{x_i}) = 0 \qquad (15.79)$$

Prin transformări elementare sistemul (15.79) devine:

$$n\,a + b \sum \frac{1}{x_i} = \sum y_i \qquad (15.80)$$

$$a \sum \frac{1}{x_i} + b \sum \frac{1}{x_i^2} = \sum \frac{y_i}{x_i}$$

Aplicînd regula lui Cramer pentru rezolvarea sistemului (15.80) rezultă formulele uzuale de calcul ale parametrilor ecuaţiri de regresie de tip hiperbolic:

$$a = \frac{\Delta_a}{\Delta} = \frac{\begin{vmatrix} \sum y_i & \sum \frac{1}{x_i} \\ \sum \frac{y_i}{x_i} & \sum \frac{1}{x_i^2} \end{vmatrix}}{\begin{vmatrix} n & \sum \frac{1}{x_i} \\ \sum \frac{1}{x_i} & \sum \frac{1}{x_i^2} \end{vmatrix}} = \frac{\sum y_i \sum \frac{1}{x_i} - \sum \frac{1}{x_i} \sum \frac{y_i}{x_i}}{n \sum \frac{1}{x_i^2} - \left(\sum \frac{1}{x_i}\right)^2} \qquad (15.81)$$

$$b = \frac{\Delta_b}{\Delta} = \frac{\begin{vmatrix} n & \sum y_i \\ \sum \frac{1}{x_i} & \sum \frac{y_i}{x_i} \end{vmatrix}}{\begin{vmatrix} n & \sum \frac{1}{x_i} \\ \sum \frac{1}{x_i} & \sum \frac{1}{x_i^2} \end{vmatrix}} = \frac{n \sum \frac{y_i}{x_i} - \sum y_i \sum \frac{1}{x_i}}{n \sum \frac{1}{x_i^2} - \left(\sum \frac{1}{x_i}\right)^2} \qquad (15.82)$$

Formulele parametrilor ecuaţiei de regresie de tip hiperbolic (15.81) şi (15.82) se utilizează în practica statistică atunci cînd valorile variabilelor corelate X şi Y se prezintă sub forma a două şiruri paralele. Calculele preliminare se prezintă sistematizat sub forma tabelului 15.7.

Dacă perechile de valori corelate x_i, y_i sînt ponderate cu frecvenţa n_i, atunci sistemul de ecuaţii normale (15.80) devine:

$$\begin{aligned} a\sum n_i + b\sum \frac{n_i}{x_i} &= \sum y_i n_i \\ a\sum \frac{n_i}{x_i} + b\sum \frac{n_i}{x_i^2} &= \sum \frac{y_i n_i}{x_i} \end{aligned} \qquad (15.83)$$

Rezolvînd sistemul (15.83) rezultă:

Tabelul 15.7

Nr. crt.	X	Y	$\frac{1}{x_i}$	$\frac{1}{x_i^2}$	$\frac{y_i}{x_i}$	y_i^2
0	1	2	3	4	5	6
1	x_1	y_1	$\frac{1}{x_1}$	$\frac{1}{x_1^2}$	$\frac{y_1}{x_1}$	y_1^2
2	x_2	y_2	$\frac{1}{x_2}$	$\frac{1}{x_2^2}$	$\frac{y_2}{x_2}$	y_2^2
. . .						
i	x_i	y_i	$\frac{1}{x_i}$	$\frac{1}{x_i^2}$	$\frac{y_i}{x_i}$	y_i^2
. . .						
n	x_n	y_n	$\frac{1}{x_n}$	$\frac{1}{x_n^2}$	$\frac{y_n}{x_n}$	y_n^2
TOTAL	$\sum x_i$	$\sum y_i$	$\sum \frac{1}{x_i}$	$\sum \frac{1}{x_i^2}$	$\sum \frac{y_i}{x_i}$	$\sum y_i^2$

$$a = \frac{\sum y_i n_i \sum \frac{n_i}{x_i} - \sum \frac{n_i}{x_i^2} \sum \frac{y_i n_i}{x_i}}{\sum n_i \sum \frac{n_i}{x_i^2} - \left(\sum \frac{n_i}{x_i}\right)^2} \qquad (15.84)$$

$$b = \frac{\sum n_i \sum \frac{y_i n_i}{x_i} - \sum \frac{n_i}{x_i} \sum y_i n_i}{\sum n_i \sum \frac{n_i}{x_i^2} - \left(\sum \frac{n_i}{x_i}\right)^2} \qquad (15.85)$$

Calculul sumelor ce apar în formulele (15.84) şi (15.85) se poate sistematiza cu ajutorul tabelului 15.8.

Dacă datele iniţiale cu privire la variabilele X şi Y se prezintă sub forma unui tabel de corelaţie, atunci sistemul

(15.80) devine:

Tabelul 15.8

Nr. crt.	X	Y	n_i	$\frac{1}{x_i}$	$\frac{n_i}{x_i}$	$\frac{1}{x_i^2}$	$\frac{n_i}{x_i^2}$	$y_i n_i$	$\frac{y_i n_i}{x_i}$	$y_i^2 n_i$
0	1	2	3	4	5	6	7	8	9	10
1	x_1	y_1	n_1	$\frac{1}{x_1}$	$\frac{n_1}{x_1}$	$\frac{1}{x_1^2}$	$\frac{n_1}{x_1^2}$	$y_1 n_1$	$\frac{y_1 n_1}{x_1}$	$y_1^2 n_1$
2	x_2	y_2	n_2	$\frac{1}{x_2}$	$\frac{n_2}{x_2}$	$\frac{1}{x_2^2}$	$\frac{n_2}{x_2^2}$	$y_2 n_2$	$\frac{y_2 n_2}{x_2}$	$y_2^2 n_2$
⋮ i	x_i	y_i	n_i	$\frac{1}{x_i}$	$\frac{n_i}{x_i}$	$\frac{1}{x_i^2}$	$\frac{n_i}{x_i^2}$	$y_i n_i$	$\frac{y_i n_i}{x_i}$	$y_i^2 n_i$
⋮ k	x_k	y_k	n_k	$\frac{1}{x_k}$	$\frac{n_k}{x_k}$	$\frac{1}{x_k^2}$	$\frac{n_k}{x_k^2}$	$y_k n_k$	$\frac{y_k n_k}{x_k}$	$y_k^2 n_k$
TOTAL	-	-	$\sum n_i$	-	$\sum \frac{n_i}{x_i}$	-	$\sum \frac{n_i}{x_i^2}$	$\sum y_i n_i$	$\sum \frac{y_i n_i}{x_i}$	$\sum y_i^2 n_i$

$$a \sum\sum n_{ij} + b \sum \frac{n_{i\cdot}}{x_i} = \sum y_j n_{\cdot j} \qquad (15.86)$$

$$a \sum \frac{n_{i\cdot}}{x_i} + b \sum \frac{n_{i\cdot}}{x_i^2} = \sum\sum \frac{y_i}{x_i} n_{ij}$$

Prin rezolvarea sistemului (15.86) rezultă formulele usuale de calcul al parametrilor ecuației de regresie de tip hiperbolic pe baza unui tabel de corelație:

$$a = \frac{\sum y_j n_{\cdot j} \sum \frac{n_{i\cdot}}{x_i^2} - \sum \frac{n_{i\cdot}}{x_i} \sum \frac{y_i}{x_i} n_{ij}}{\sum\sum n_{ij} \sum \frac{n_{i\cdot}}{x_i^2} - \left(\sum \frac{n_{i\cdot}}{x_i}\right)^2} \qquad (15.87)$$

$$b = \frac{\sum\sum n_{ij} \sum \frac{y_i}{x_i} n_{ij} - \sum \frac{n_{i\cdot}}{x_i} \sum y_j n_{\cdot j}}{\sum\sum n_{ij} \sum \frac{n_{i\cdot}}{x_i^2} - \left(\sum \frac{n_{i\cdot}}{x_i}\right)^2} \qquad (15.88)$$

Scanned with OKEN Scanner

(15.80) devine:

Tabelul 15.8

Nr. crt.	X	Y	n_i	$\frac{1}{x_i}$	$\frac{n_i}{x_i}$	$\frac{1}{x_i^2}$	$\frac{n_i}{x_i^2}$	$y_i n_i$	$\frac{y_i n_i}{x_i}$	$y_i^2 n_i$
0	1	2	3	4	5	6	7	8	9	10
1	x_1	y_1	n_1	$\frac{1}{x_1}$	$\frac{n_1}{x_1}$	$\frac{1}{x_1^2}$	$\frac{n_1}{x_1^2}$	$y_1 n_1$	$\frac{y_1 n_1}{x_1}$	$y_1^2 n_1$
2	x_2	y_2	n_2	$\frac{1}{x_2}$	$\frac{n_2}{x_2}$	$\frac{1}{x_2^2}$	$\frac{n_2}{x_2^2}$	$y_2 n_2$	$\frac{y_2 n_2}{x_2}$	$y_2^2 n_2$
⋮ i	x_i	y_i	n_i	$\frac{1}{x_i}$	$\frac{n_i}{x_i}$	$\frac{1}{x_i^2}$	$\frac{n_i}{x_i^2}$	$y_i n_i$	$\frac{y_i n_i}{x_i}$	$y_i^2 n_i$
⋮ k	x_k	y_k	n_k	$\frac{1}{x_k}$	$\frac{n_k}{x_k}$	$\frac{1}{x_k^2}$	$\frac{n_k}{x_k^2}$	$y_k n_k$	$\frac{y_k n_k}{x_k}$	$y_k^2 n_k$
TOTAL	-	-	$\sum n_i$	-	$\sum \frac{n_i}{x_i}$	-	$\sum \frac{n_i}{x_i^2}$	$\sum y_i n_i$	$\sum \frac{y_i n_i}{x_i}$	$\sum y_i^2 n_i$

$$a \sum\sum n_{ij} + b \sum \frac{n_{i\cdot}}{x_i} = \sum y_j n_{\cdot j} \qquad (15.86)$$

$$a \sum \frac{n_{i\cdot}}{x_i} + b \sum \frac{n_{i\cdot}}{x_i^2} = \sum\sum \frac{y_i}{x_i} n_{ij}$$

Prin rezolvarea sistemului (15.86) rezultă formulele usuale de calcul al parametrilor ecuației de regresie de tip hiperbolic pe baza unui tabel de corelație:

$$a = \frac{\sum y_j n_{\cdot j} \sum \frac{n_{i\cdot}}{x_i^2} - \sum \frac{n_{i\cdot}}{x_i} \sum \frac{y_i}{x_i} n_{ij}}{\sum\sum n_{ij} \sum \frac{n_{i\cdot}}{x_i^2} - \left(\sum \frac{n_{i\cdot}}{x_i}\right)^2} \qquad (15.87)$$

$$b = \frac{\sum\sum n_{ij} \sum \frac{y_i}{x_i} n_{ij} - \sum \frac{n_{i\cdot}}{x_i} \sum y_j n_{\cdot j}}{\sum\sum n_{ij} \sum \frac{n_{i\cdot}}{x_i^2} - \left(\sum \frac{n_{i\cdot}}{x_i}\right)^2} \qquad (15.88)$$

Scanned with OKEN Scanner

Tabelul 15.9

Nr. crt.	Y \ X	x_1	x_2	...	x_i	...	x_s	Total $(n_{\cdot j})$	$y_j n_{\cdot j}$	$\sum_i \frac{y_j}{x_i} n_{ij}$	$y_j^2 n_{\cdot j}$
0	A	1	2	...	i	...	s	s+1	s+2	s+3	s+4
1	y_1	n_{11}	n_{21}	...	n_{i1}	...	n_{s1}	$n_{\cdot 1}$	$y_1 n_{\cdot 1}$	$\sum_i \frac{y_1}{x_i} n_{i1}$	$y_1^2 n_{\cdot 1}$
2	y_2	n_{12}	n_{22}	...	n_{i2}	...	n_{s2}	$n_{\cdot 2}$	$y_2 n_{\cdot 2}$	$\sum_i \frac{y_2}{x_i} n_{i2}$	$y_2^2 n_{\cdot 2}$
:											
j	y_j	n_{1j}	n_{2j}	...	n_{ij}	...	n_{sj}	$n_{\cdot j}$	$y_j n_{\cdot j}$	$\sum_i \frac{y_j}{x_i} n_{ij}$	$y_j^2 n_{\cdot j}$
:											
k	y_k	n_{1k}	n_{2k}	...	n_{ik}	...	n_{sk}	$n_{\cdot k}$	$y_k n_{\cdot k}$	$\sum_i \frac{y_k}{x_i} n_{ik}$	$y_k^2 n_{\cdot k}$
k+1	Total	$n_{1\cdot}$	$n_{2\cdot}$	...	$n_{i\cdot}$	...	$n_{s\cdot}$	$n_{\cdot\cdot}$	$\sum_j y_j n_{\cdot j}$	$\sum_i \sum_j \frac{y_j}{x_i} n_{ij}$	$\sum_j y_j^2 n_{\cdot j}$
k+2	$\frac{n_{1\cdot}}{x_1}$	$\frac{n_{1\cdot}}{x_1}$	$\frac{n_{2\cdot}}{x_2}$	...	$\frac{n_{i\cdot}}{x_i}$	...	$\frac{n_{s\cdot}}{x_s}$	$\sum \frac{n_{i\cdot}}{x_i}$			
k+3	$\frac{n_{1\cdot}}{x_1^2}$	$\frac{n_{1\cdot}}{x_1^2}$	$\frac{n_{2\cdot}}{x_2^2}$	...	$\frac{n_{i\cdot}}{x_i^2}$	...	$\frac{y_1}{x_s^2}$	$\sum \frac{n_{i\cdot}}{x_i^2}$			
k+4	$\sum_j \frac{y_1}{x_1} n_{1\cdot}$	$\sum_j \frac{y_1}{x_1} n_{1j}$	$\sum_j \frac{y_1}{x_2} n_{2j}$	...	$\sum_j \frac{y_1}{x_1} n_{ij}$	...	$\sum_j \frac{y_1}{x_s} n_{sj}$	$\sum_i \sum_j \frac{y_1}{x_i} n_{ij}$			
k+5	y_{x_1}	y_{x_1}	y_{x_2}	...	y_{x_i}	...	y_{x_s}	-			

Scanned with OKEN Scanner

Calculul sumelor,ce apar în formulele (15.87) şi (15.88), se prezintă în tabelul 15.9.

Pentru ilustrarea modului de calcul şi analiză statistică a legăturilor simple curbilinii, să considerăm dependenţa dintre volumul desfacerilor de mărfuri în 44 de unităţi comerciale şi nivelul relativ al cheltuielilor de circulaţie. Datele iniţiale se prezintă în tabelul 15.10.

Tabelul 15.10. Vînzările de mărfuri şi nivelul relativ al cheltuielilor de circulaţie

Nivelul relativ al cheltuielilor de circulaţie (%) \ Volumul vînzărilor de mărfuri (milioane lei)	Pînă la 4	4-8	8-12	12-16	16-20	20 şi peste	TOTAL
Pînă la 3,6	-	-	-	-	4	5	9
3,6-4,2	-	-	5	3	1	2	11
4,2-4,8	-	5	3	-	-	-	8
4,8-5,4	1	4	-	-	-	-	5
5,4-6,0	3	2	-	-	-	-	5
6,0 şi peste	6	-	-	-	-	-	6
T O T A L	10	11	8	3	5	7	44

Date convenţionale

Este ştiut faptul că unităţile comerciale mici au un nivel mai ridicat al cheltuielilor de circulaţie decît unităţile ce realizează un volum mai mare al vînzărilor de mărfuri. Pe măsură ce creşte volumul desfacerilor de mărfuri se reduce în mod continuu nivelul relativ al cheltuielilor de circulaţie. La început reducerea este mai mare, iar pe măsura creşterii volumului vînzărilor, această reducere devine din ce în ce mai mică. Acest fapt apare mai evident dacă vom observa modul de dispersare al frecvenţelor în cadrul tabelului 15.11.Deci legătura dintre variabile se apropie de modelul unei hiperbole.Un alt argument în favoarea ipotezei dependenţei hiperbolice va fi dat de corelograma din fig.15.9.

Scanned with OKEN Scanner

Tabelul 15.11. Calculele preliminare pentru determinarea indicatorilor de regresie şi corelaţie

Nr. crt.	X / Y	x_i / y_j	- 4	4 - 8	8 - 12	12 - 16	16-20	20 şi peste	TOTAL ($n_{.j}$)	$y_j n_{.j}$	$y_j^2 n_{.j}$	$\sum_i \frac{y_j}{x_i} n_{ij}$	Medii condiţionate $\bar{y}/y_i$
0	1		2	3	4	5	6	7	8	9	10	11	12
1	-3,6	3,3	-	-	-	-	4	5	9	29,7	98,01	1,483332	20,2
2	3,6-4,2	3,9	-	-	5	3	1	2	11	42,9	167,31	3,356923	14,0
3	4,2-4,8	4,5	-	5	3	-	-	-	8	36,0	162,00	5,100000	7,5
4	4,8-5,4	5,1	1	4	-	-	-	-	5	25,5	130,05	5,950000	5,2
5	5,4-6,0	5,7	3	2	-	-	-	-	5	28,5	162,45	10,450000	3,6
6	6,0 şi peste	6,3	6	-	-	-	-	-	6	37,8	238,14	18,900000	2,0
7	Total ($n_{i.}$)		10	11	8	3	5	7	44	200,4	957,96	45,240255	-
9	$\frac{n_{i.}}{x_i}$		5,0000	1,8333	0,8000	0,2143	0,2778	0,3182	8,4436				
9	$\frac{n_{i.}}{x_i^2}$		2,5000	0,3056	0,0800	0,0153	0,0154	0,0145	2,9308				
10	$\sum_j \frac{y_j}{x_i} n_{ij}$		30,00000	9,050000	3,300000	0,835713	0,949999	1,104544	45,240255				
11	y_x		6,15	4,42	4,08	3,93	3,85	3,80	-				
12	Medii condiţionate $\bar{y}/x_i$		6,0	4,94	4,13	3,9	3,42	3,47	-				

Scanned with OKEN Scanner

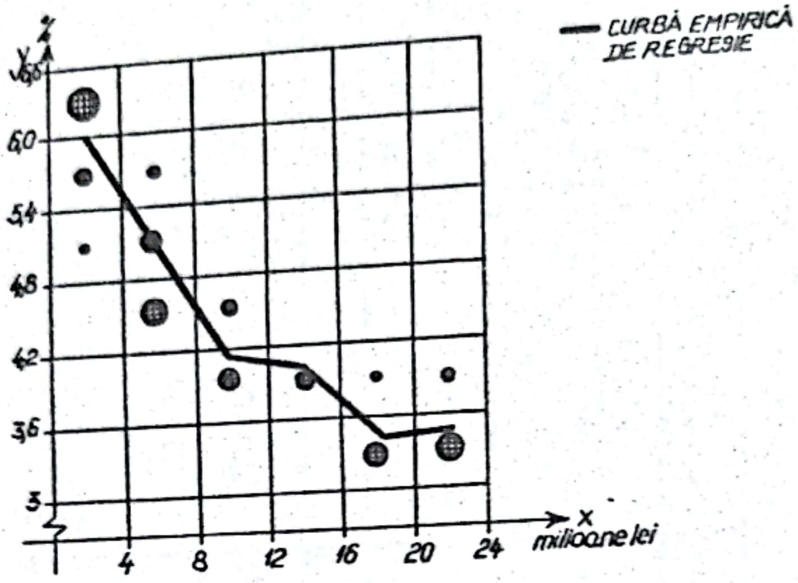


Fig. 15.9. CORELAȚIA DINTRE NIVELUL RELATIV AL CHELTUIELILOR DE CIRCULAȚIE ȘI VOLUMUL VÎNZĂRILOR DE MĂRFURI

Un alt argument în favoarea ipotezei dependenței hiperbolice dintre nivelul relativ al cheltuielilor de circulație și volumul vînzărilor de mărfuri se obține calculînd mediile condiționate ale variabilei Y în funcție de diferite valori ale lui X .

Parametrii ecuației de regresie $y = a + \frac{b}{x}$ se determină cu ajutorul formulelor (15.87) și (15.88):

$$a = \frac{\sum y_j n_{\cdot j} \sum \frac{n_{i\cdot}}{x_i^2} - \sum \frac{n_{i\cdot}}{x_i} \sum \frac{y_i}{x_i} n_{ij}}{\sum\sum n_{ij} \sum \frac{n_{i\cdot}}{x_i^2} - \left(\sum \frac{n_{i\cdot}}{x_i}\right)^2} = \frac{200,4.2,9308 - 8,4436.45,23946}{44.2,9308 - 8,4436^2} =$$

$$= 3,56$$

$$b = \frac{\sum\sum n_{ij} \sum\sum \frac{y_i}{x_i} n_{ij} - \sum \frac{n_{i\cdot}}{x_i} \sum y_j n_{\cdot j}}{\sum\sum n_{ij} \sum \frac{n_{i\cdot}}{x_i^2} - \left(\sum \frac{n_{i\cdot}}{x_i}\right)^2} = \frac{44.45,23946 - 8,4436.200,4}{44.2,9308 - 8,4436^2} =$$

$$= 5,17576 \text{ milioane lei}$$

Corelația dintre nivelul relativ al cheltuielilor de circulație și volumul vînzărilor de mărfuri este descrisă de ecuația de regresie:

$$y = 3,56 + \frac{5,18}{x}$$

Ecuaţia de regresie ne arată că, în condiţiile celor 44 de unităţi comerciale luate în studiu, la 100 lei vînzări revin 3,56 lei cheltuieli variabile şi 5,18 milioane lei cheltuieli convenţional constante.

Nivelul relativ al cheltuielilor de circulaţie corespunzător diferitelor valori ale volumului vînzărilor de mărfuri se prezintă în rîndul 11 al tabelului 15.11.

Intensitatea legăturii dintre variabile, în cazul unei legături simple curbilinii de tip hiperbolic, se poate determina cu ajutorul formulei generale a raportului de corelaţie:

$$\eta = \sqrt{\frac{\sigma^2_{y_x}}{\sigma^2_y}} = \sqrt{1 - \frac{\sum(y_i - y_x)^2}{\sum(y_i - \bar{y})^2}}$$

Sistematizarea calculelor pentru aplicarea acestei formule se prezintă în tabelul 15.9.

Ca şi în cazul regresiei parabolice, în cazul unei hiperbole se poate stabili o formulă de calcul simplificat a raportului de corelaţie η .

După cum am arătat, în cazul legăturii curbilinii de tip parabolic şi în cazul hiperbolei are loc relaţia:

$$\sum(y_i - y_x)^2 = \sum(y_i - y_x)(y_i - a - \frac{b}{x})$$

$$\sum(y_i - y_x)^2 = \sum(y_i - y_x)\, y_i - a\sum(y_i - y_x) - b\sum(y_i - y_x)\frac{1}{x}$$

Dar

$$a\sum(y_i - y_x) = a\sum(y_i - a - \frac{b}{x_i}) = a\,(\sum y_i - n\,a - b\sum\frac{1}{x}) = 0$$

şi

$$b\sum(y_i - y_x)\frac{1}{x} = b\sum(y_i - a - \frac{b}{x_i})\frac{1}{x_i} = b\,(\sum\frac{y_i}{x_i} - a\sum\frac{1}{x_i} - b\sum\frac{1}{x_i^2}) = 0$$

Rezultă

$$\sum(y_i - y_x)^2 = \sum(y_i - y_x)\, y_i$$

Substituind pe y_x cu expresia lui, în membrul al doilea al formulei anterioare, rezultă:

$$\sum(y_i - a - \frac{b}{x_i})y_i = \sum y_i^2 - a\sum y_i - b\sum\frac{y_i}{x_i} \qquad (15.89)$$

Expresia de la numitorul raportului de corelaţie se poate scrie sub forma:

$$\sum(y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n} \qquad (15.90)$$

In baza relaţiei (15.90) şi (15.89), formula raportului de corelaţie devine:

$$\eta = \sqrt{\frac{a\sum y_i + b\sum \frac{y_i}{x_i} - \frac{(\sum y_i)^2}{n}}{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}} \qquad (15.91)$$

Dacă perechile de variante x_i, y_i ale celor două variabile sînt ponderate cu frecvenţa n_i, atunci, formula de calcul a raportului de corelaţie, pentru dependenţa de tip parabolic, devine:

$$\eta = \sqrt{\frac{a\sum y_i n_i + b\sum \frac{y_i}{x_i} n_i - \frac{(\sum y_i n_i)^2}{\sum n_i}}{\sum y_i^2 n_i - \frac{(\sum y_i n_i)^2}{\sum n_i}}} \qquad (15.92)$$

Iar în cazul unui tabel de corelaţie η va fi:

$$\eta = \sqrt{\frac{a\sum y_j n_{\cdot j} + b\sum\sum \frac{y_i}{x_i} n_{ij} - \frac{(\sum y_i n_{\cdot i})^2}{\sum\sum n_{ij}}}{\sum y_j^2 n_{\cdot j} - \frac{(\sum y_i n_{\cdot i})^2}{\sum\sum n_{ij}}}} \qquad (15.93)$$

Toate sumele ce apar în formula (15.93) sînt calculate în tabelul 15.11 pentru exemplul anterior. Substituind aceste valori în formula (15.93), rezultă:

$$\eta = \sqrt{\frac{3,56.200,4+5,18.45,24 - \frac{(200,4)^2}{44}}{957,96 - \frac{200,4^2}{44}}} = \sqrt{\frac{35,0363}{45,229}} = \sqrt{0,774642} = 0,88$$

$\eta = 0,88$ înseamnă că între volumul vînzărilor de mărfuri şi nivelul relativ al cheltuielilor de circulaţie există o corelaţie strînsă.

In cazul unor legături curbilinii, calculul şi analiza indicatorilor de regresie şi corelaţie poate deveni mai complicat. Pentru a evita unele calcule mai complexe şi mai greu de interpretat, în diferite situaţii concrete, se recurge la un "artificiu de calcul" cunoscut în literatura de specialitate sub denumirea de liniarizarea legăturilor curbilinii. Prin aplicarea acestui procedeu, funcţii complicate de regresie se reduc la o formă liniară, uşor de calculat şi apoi indicatorii de regresie şi corelaţie se determină după aceleaşi procedee ca şi în cazul legăturilor liniare. Prin liniarizare, după cum vom arăta în continuare, nu se schimbă nici forma reală a legăturilor curbilinii, nici esenţa indicatorilor de regresie şi corelaţie, ci doar se simplifică calculele.

Ecuaţia hiperbolei (15.75) se liniarizează astfel:

$$y' = a + b\frac{1}{x'}$$

Substituind pe $x = \frac{1}{x'}$, rezultă:

$$y = a + b\,x$$

Intensitatea corelaţiei în acest se poate determina, ca în cazul oricărei legături liniare, cu ajutorul coeficientului de corelaţie, astfel:

$$r_{\frac{1}{x},y} = \frac{n\sum\frac{1}{x_i}y_i - \sum\frac{1}{x_i}\sum y_i}{\sqrt{\left[n\sum\left(\frac{1}{x_i}\right)^2 - \left(\sum\frac{1}{x_i}\right)^2\right]\left[n\sum y_i^2 - \left(\sum y_i\right)^2\right]}} \qquad (15.94)$$

Remarcăm faptul că, atît coeficientul de corelaţie (15.94) cît şi raportul de corelaţie, calculaţi pentru aceleaşi date vor da acelaşi rezultat. Am demonstrat, de altfel, că în cazul legăturilor liniare (şi a celor liniarizate), raportul de corelaţie este identic cu coeficientul de corelaţie.

Coeficientul de corelaţie în cazul legăturilor curbilinii liniarizate se prezintă într-o formă mai convenabilă de aplicare în practică, se calculează şi se interpretează mai uşor. La fel ca şi raportul de corelaţie, oferă posibilitatea măsurării într-o formă unitară a intensităţii legăturii dintre variabile, indiferent dacă sînt liniare sau curbilinii. Măreşte în acest fel posibilităţile de comparabilitate a intensităţii legăturii de la un tip de ecuaţie de regresie la altul.

Desigur, formula coeficientului de corelaţie va avea expresii diferite în funcţie de modul de prezentare a variabilelor corelate. Astfel, dacă perechile de valori ale celor două variabile sînt neponderate, atunci formula de calcul este cea din (15.94). Dacă perechile de valori x_i, y_i sînt ponderate cu frecvenţa n_i, atunci coeficientul de corelaţie, pentru o legătură hiperbolică liniarizată, va fi dat de formula:

$$r_{\frac{1}{x},y} = \frac{\sum n_i \sum \frac{n_i y_i}{x_i} - \sum \frac{n_i}{x_i} \sum y_i n_i}{\sqrt{\left[\sum n_i \sum \frac{n_i}{x_i^2} - \left(\sum \frac{n_i}{x_i}\right)^2\right]\left[\sum n_i \sum y_i^2 n_i - \left(\sum y_i n_i\right)^2\right]}} \qquad (15.95)$$

Dacă variabilele corelate hiperbolic se prezintă sub forma unui tabel de corelaţie cu dublă intrare, atunci coeficientul de corelaţie va avea formula:

$$r_{\frac{1}{x},y} = \frac{\sum\sum n_{ij} \sum\sum \frac{n_{ij} y_j}{x_i} - \sum \frac{n_{i.}}{x_i} \sum n_{.j} y_j}{\sqrt{\left[\sum\sum n_{ij} \sum \frac{n_{i.}}{x_i^2} - \left(\sum \frac{n_{i.}}{x_i}\right)^2\right]\left[\sum\sum n_{ij} \sum n_{.j} y_j^2 - \left(\sum n_{.j} y_j\right)^2\right]}} \qquad (15.96)$$

In cazul regresiei hiperbolice, date de modelul (15.77), al cărei grafic se prezintă în fig. 15.8, legătura dintre variabile poate avea expresiile:

$$y = \frac{1}{a + b x} \qquad \text{sau} \qquad \frac{1}{y} = a + bx$$

Determinarea parametrilor a şi b ai acestei funcţii cu ajutorul metodei celor mai mici pătrate se face pe baza sistemului de ecuaţii normale:

$$n\,a + b\sum x_i = \sum \frac{1}{y_i} \qquad (15.97)$$

$$a\sum x_i + b\sum x_i^2 = \sum \frac{x_i}{y_i}$$

Aplicînd regula lui Cramer, pentru rezolvarea sistemului (15.97) rezultă următoarele formule de calcul al celor doi parametri:

$$a = \frac{\sum x_i^2 \sum \frac{1}{y_i} - \sum x_i \sum \frac{x_i}{y_i}}{n\sum x_i^2 - (\sum x_i)^2} \qquad (15.98)$$

$$b = \frac{n\sum \frac{x_i}{y_i} - (\sum x_i \sum \frac{1}{y_i}}{n\sum x_i^2 - (\sum x_i)^2} \qquad (15.99)$$

Formulele (15.98) şi (15.99) sînt aplicabile pentru determinarea parametrilor ecuaţiei de regresie hiperbolică atunci cînd datele iniţiale cu privire la variaţia celor două variabile se prezintă sub forma a două şiruri paralele.

Dacă perechile de valori corelate x_i, y_i sînt ponderate cu n_i, atunci sistemul de ecuaţii normale (15.97) devine:

$$a\sum n_i + b\sum x_i n_i = \sum \frac{n_i}{y_i} \qquad (15.100)$$

$$a\sum x_i n_i + b\sum x_i^2 n_i = \sum \frac{x_i n_i}{y_i}$$

de unde

$$a = \frac{\sum x_i^2 n_i \sum \frac{n_i}{y_i} - \sum x_i n_i \sum \frac{x_i n_i}{y_i}}{\sum n_i \sum x_i^2 n_i - (\sum x_i n_i)^2} \qquad (15.101)$$

$$b = \frac{\sum n_i \sum \frac{x_i n_i}{y_i} - \sum x_i n_i \sum \frac{n_i}{y_i}}{\sum n_i \sum x_i^2 n_i - (\sum x_i n_i)^2} \qquad (15.102)$$

In cazul prezentării variabilelor empirice sub forma unui tabel de corelaţie cu dublă intrare, sistemul de ecuaţii normale necesar pentru determinarea parametrilor ecuaţiei de regresie (15.75) este de forma:

$$\begin{aligned} a \sum\sum n_{ij} + b \sum x_i n_{i\cdot} &= \sum \frac{n_{\cdot j}}{y_j} \\ a \sum x_i n_{i\cdot} + b \sum x_i^2 n_{i\cdot} &= \sum\sum \frac{x_i}{y_j} n_{ij} \end{aligned} \qquad (15.103)$$

de unde

$$a = \frac{\sum x_i^2 n_{i\cdot} \sum \frac{n_{\cdot j}}{y_j} - \sum x_i n_{i\cdot} \sum\sum \frac{x_i n_{ij}}{y_j}}{\sum\sum n_{ij} \sum x_i^2 n_{i\cdot} - (\sum x_i n_{i\cdot})^2} \qquad (15.104)$$

$$b = \frac{\sum\sum n_{ij} \sum\sum \frac{x_i}{y_j} n_{ij} - \sum x_i n_{i\cdot} \sum y_j n_{\cdot j}}{\sum\sum n_{ij} \sum x_i^2 n_{i\cdot} - (\sum x_i n_{i\cdot})^2} \qquad (15.105)$$

Legătura hiperbolică, de tipul (15.76), se liniarizează prin substituţia $y = \frac{1}{y}$, în ecuaţia $\frac{1}{y'} = a + bx$ şi se obţine:

$$y = a + b\,x$$

Expresiile coeficientului de corelaţie, în cazul legăturii hiperbolice (15.75) liniarizate, în funcţie de modul de sis-

tematizare a datelor empirice cu privire la cele două variabile corelate, se prezintă astfel:

- în primul caz:

$$r_{x,\frac{1}{y}} = \frac{n\sum\frac{x_i}{y_i} - \sum x_i \sum\frac{1}{y_i}}{\sqrt{\left[n\sum x_i^2 - (\sum x_i)^2\right]\left[n\sum\frac{1}{y_i^2} - (\sum\frac{1}{y_i})^2\right]}} \qquad (15.106)$$

- în cazul al doilea:

$$r_{x,\frac{i}{y}} = \frac{\sum n_i \sum\frac{n_i x_i}{y_i} - \sum n_i x_i \sum\frac{n_i}{y_i}}{\sqrt{\left[\sum n_i \sum n_i x_i^2 - (\sum n_i x_i)^2\right]\left[\sum n_i \sum\frac{n_i}{y_i^2} - (\sum\frac{n_i}{y_i})^2\right]}} \qquad (15.107)$$

- în cazul al treilea:

$$r_{x,\frac{i}{y}} = \frac{\sum\sum n_{ij} \sum\frac{n_{ij} x_i}{y_j} - \sum n_{i.} x_i \sum\frac{n_{i.}}{y_j}}{\sqrt{\left[\sum\sum n_{ij} \sum n_{i.} x_i^2 - (\sum n_{i.} x_i)^2\right]\left[\sum\sum n_{ij} \sum\frac{n_{.j}}{y_j^2} - (\sum\frac{n_{.j}}{y_j})^2\right]}} \qquad (15.108)$$

În condiţiile legăturii hiperbolice (15.77) dată de ecuaţia:

$$y' = \frac{x'}{a'x' + b} = \frac{1}{a' + \frac{b'}{x}}$$

substituind pe $y = \frac{1}{y'}$ şi pe $x = \frac{1}{x'}$, se obţine:

$$y = a' + b'x$$

Coeficientul de corelaţie pentru această legătură hiperbolică liniarizată, în funcţie de modul de sistematizare a datelor iniţiale cu privire la variabilele corelate, se prezintă astfel:

- în primul caz:

$$r_{\frac{1}{x},\frac{1}{y}} = \frac{n\sum\frac{1}{x_i}\,\frac{1}{y_i} - \sum\frac{1}{x_i}\sum\frac{1}{y_i}}{\sqrt{\left[n\sum\frac{1}{x_i^2} - \left(\sum\frac{1}{x_i}\right)^2\right]\left[n\sum\frac{1}{y_i^2} - \left(\sum\frac{1}{y_i}\right)^2\right]}} \qquad (15.109)$$

- în cazul al doilea:

$$r_{\frac{1}{x},\frac{1}{y}} = \frac{\sum n_i \sum\frac{n_i}{x_i y_i} - \sum\frac{n_i}{x_i}\sum\frac{n_i}{y_i}}{\sqrt{\left[\sum n_i\sum\frac{n_i}{x_i^2} - \left(\sum\frac{n_i}{x_i}\right)^2\right]\left[\sum n_i\sum\frac{n_i}{y_i^2} - \left(\sum\frac{n_i}{y_i}\right)^2\right]}} \qquad (15.110)$$

- în cazul al treilea:

$$r_{\frac{1}{x},\frac{1}{y}} = \frac{\sum\sum n_{ij}\sum\sum\frac{n_{ij}}{x_i y_j} - \sum\frac{n_{i\cdot}}{x_i}\sum\frac{n_{\cdot j}}{y_j}}{\sqrt{\left[\sum\sum n_{ij}\sum\frac{n_{i\cdot}}{x_i^2} - \left(\sum\frac{n_{i\cdot}}{x_i}\right)^2\right]\left[\sum\sum n_{ij}\sum\frac{n_{\cdot j}}{y_j^2} - \left(\sum\frac{n_{\cdot j}}{y_j}\right)^2\right]}} \qquad (15.111)$$

15.1.3. Legături după modelul unei funcţii putere

Regresia după modelul unei funcţii putere este descrisă de ecuaţia:

$$y = a\,x^b \quad \text{sau} \quad \log y = \log a + b \log x \qquad (15.112)$$

Expresiile geometrice, descrise de ecuaţia (15.112), se prezintă în fig. 15.10.

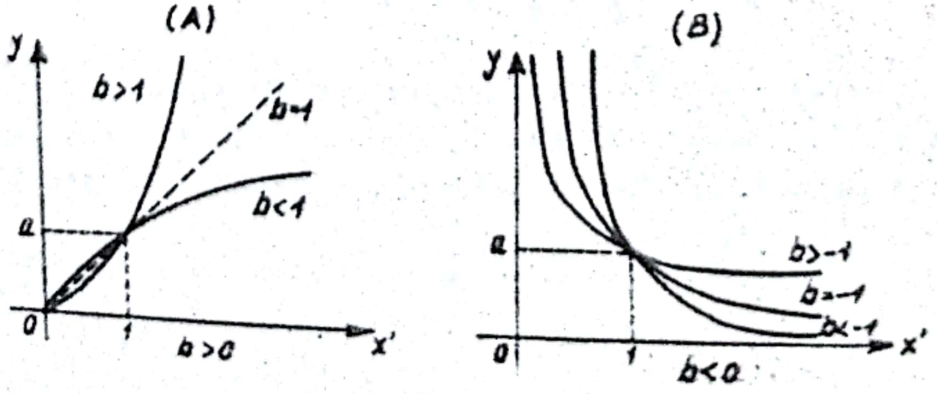


Fig. 15.10 FUNCTIA DE REGRESIE TIP PUTERE $y = ax^b$

Observăm din fig. 15.10, că funcţia tip putere trece prin punctul (1,a) oricare ar fi valorile parametrilor a şi b . In cazul în care coeficientul de regresie b este pozitiv (fig.15.10 A) şi cuprins între 0 şi 1,curba descrisă de funcţia de putere trece prin originea axelor de coordonate, este crescătoare şi convexă cînd $b < 1$, iar pentru $b > 1$ curba este tot crescătoare dar concavă şi cînd $b = 1$ curba se transformă într-o dreaptă ce trece prin origine. Cînd b este negativ curbele trec prin punctul de coordonate (1,a) şi au ca asimptote axele de coordonate (fig. 15.10 B).

In cazul funcţiilor putere, legătura dintre variabile se manifestă astfel: unor creşteri relative (procentuale) egale ale variabilei X le corespund creşteri relative (procentuale) egale pentru variabila Y .

Sistemul de ecuaţii normale,obţinut pe baza metodei celor mai mici pătrate, în cazul unei funcţii putere:

$$n \log a + b \sum \log x_i = \sum \log y_i \qquad (15.113)$$

$$\log a \sum \log x_i + b \sum (\log x_i)^2 = \sum (\log x_i)(\log y_i)$$

de unde rezultă:

$$\log a = \frac{\sum \log y_i \sum (\log x_i)^2 - \sum (\log x_i) \sum (\log x_i)(\log y_i)}{n \sum (\log x_i)^2 - (\sum \log x_i)^2} \qquad (15.114)$$

$$b = \frac{n \sum (\log x_i)(\log y_i) - \sum \log x_i \sum \log y_i}{n \sum (\log x_i)^2 - (\sum \log x_i)^2} \qquad (15.115)$$

Dacă perechile de valori empirice corelate x_i, y_i sînt ponderate cu frecvenţa n_i, atunci sistemul de ecuaţii normale (15.113), devine:

$$\log a \sum n_i + b \sum n_i \log x_i = \sum n_i \log y_i \qquad (15.116)$$

$$\log a \sum n_i \log x_i + b \sum n_i (\log x_i)^2 = \sum n_i (\log x_i)(\log y_i)$$

Scanned with OKEN Scanner

Prin rezolvarea sistemului (15.116) rezultă:

$$\log a = \frac{\sum n_i \log y_i \sum n_i (\log x_i)^2 - \sum n_i \log x_i \sum n_i (\log x_i)(\log y_i)}{\sum n_i \sum n_i (\log x_i)^2 - (\sum n_i \log x_i)^2} \qquad (15.117)$$

$$b = \frac{\sum n_i \sum n_i (\log x_i)(\log y_i) - \sum n_i \log x_i \sum n_i \log y_i}{\sum n_i \sum n_i (\log x_i)^2 - (\sum n_i \log x_i)^2} \qquad (15.118)$$

In cazul prezentării datelor inițiale, cu privire la variația fenomenelor corelate, sub formă de tabel de corelație cu dublă intrare, sistemul de ecuații normale va fi de forma:

$$\log a \sum\sum n_{ij} + b \sum n_{i.} \log x_i = \sum n_{.j} \log y_j \qquad (15.119)$$

$$\log a \sum n_i . \log x_i + b \sum n_{i.} (\log x_i)^2 = \sum\sum n_{ij} (\log x_i)(\log y_j)$$

Din rezolvarea sistemului (15.119) rezultă formulele uzuale de calcul ale parametrilor ecuației de regresie de tip funcție putere (15.112):

$$\log a = \frac{\sum n_{.j} \log y_j \sum n_{i.} (\log x_i)^2 - \sum n_{i.} \log x_i \sum\sum n_{ij} (\log x_i)(\log y_j)}{\sum\sum n_{ij} \sum n_{i.} (\log x_i)^2 - (\sum n_{i.} \log x_i)^2} \qquad (15.120)$$

$$b = \frac{\sum\sum n_{ij} \sum\sum n_{ij} (\log x_i)(\log y_i) - \sum n_{i.} \log x_i \sum n_{.j} \log y_i}{\sum\sum n_{ij} \sum n_{i.} (\log x_i)^2 - (\sum n_{i.} \log x_i)^2} \qquad (15.121)$$

Liniarizarea funcției putere $y = a x'^b$ se obține făcînd substituția $y = \log y'$ și $x = \log x'$ în ecuația funcției putere.

Coeficientul de corelație calculat pe baza funcției putere liniarizate, în cele trei cazuri posibile de sistematizare a valorilor variabile empirice, se prezintă astfel:

- în primul caz:

$$r_{\log x, \log y} = \frac{n \sum (\log x_i)(\log y_i) - \sum \log x_i \sum \log y_i}{\sqrt{\left[n \sum (\log x_i)^2 - (\sum \log x_i)^2\right]\left[n \sum (\log y_i)^2 - (\sum \log y_i)^2\right]}} \qquad (15.122)$$

- în cazul al doilea: $r_{\log x,\log y} =$

$$= \frac{\sum n_i \sum n_i (\log x_i)(\log y_i) - \sum n_i \log x_i \sum n_i \log y_i}{\sqrt{\left[\sum n_i \sum n_i (\log x_i)^2 - (\sum n_i \log x_i)^2\right]\left[\sum n_i \sum n_i (\log y_i)^2 - (\sum n_i \log y_i)^2\right]}} \quad (15.123)$$

- în cazul al treilea: $r_{\log x,\log y} =$

$$= \frac{\sum\sum n_{ij} \sum\sum n_{ij} (\log x_i)(\log y_i) - \sum n_{i.} \log x_i \sum n_{.j} \log y_j}{\sqrt{\left[\sum\sum n_{ij} \sum n_{i.} (\log x_i)^2 - (\sum n_{i.} \log x_i)^2\right]\left[\sum\sum n_{ij} \sum n_{.j} (\log y_j)^2 - (\sum n_{.j} \log y_j)^2\right]}} \quad (15.124)$$

Să considerăm pentru exemplificarea calculului şi analizei indicatorilor de regresie şi corelaţie în cazul unei funcţii de tip putere, legătura dintre distanţele faţă de soare şi perioadele de revoluţie ale planetelor sistemului nostru solar (Tabelul 15.12).

Tabelul 15.12. Distanţele planetelor sistemului solar faţă de soare şi perioadele lor de revoluţie

Planeta	Distanţa faţă de soare (în unităţi astronomice)	Perioada de revoluţie (ani)
Mercur	0,39	0,24
Venus	0,72	0,62
Pământ	1,00	1,00
Marte	1,52	1,88
Jupiter	5,20	11,90
Saturn	9,54	29,50
Uranus	19,20	84,00
Neptun	30,10	165,00
Pluto	39,50	248,00

Din punct de vedere astronomic legătura dintre distanţele planetelor faţă de soare şi perioadele lor de revoluţie este evidentă. Pe măsură ce creşte distanţa planetelor faţă de soare, creşte şi perioada lor de revoluţie.

Pentru a ne forma o imagine asupra formei legăturii dintre distanţa planetelor faţă de soare (variabila independentă X) şi perioada de revoluţie (considerată variabilă rezultativă Y) vom recurge la reprezentarea grafică a acestei corelaţii (Fig.

15.11). Din fig. 15.11 se observă că legătura dintre variabile se apropie de forma unei curbe logaritmice (de tip putere), care poate fi descrisă de ecuația: $y = a\,x^b$. Aceasta înseamnă că, în urma liniarizării acestei ecuații prin logaritmare, noua ecuație $\log y = \log a + b \log x$ reprezintă o dreaptă care depinde de parametrii $\log a$ și b . Legătura liniarizată s-a stabilit între $\log x_i$ și $\log y_i$. Pentru reprezentarea grafică a acestei legături se utilizează o hîrtie dublulogaritmică (fig. 15.12).

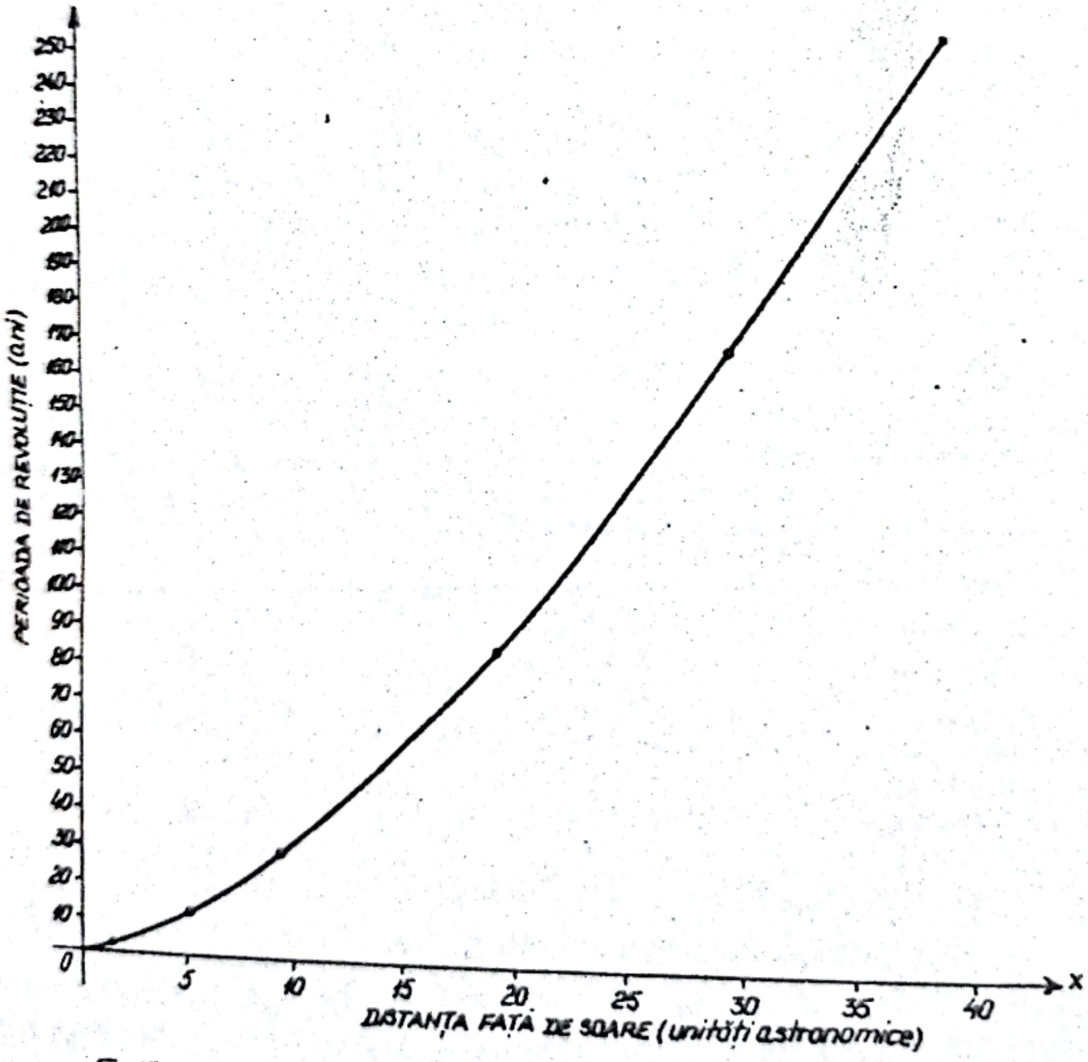


Fig.15.11 LEGĂTURA DINTRE DISTANȚA PLANETELOR FAȚĂ DE SOARE ȘI PERIOADA LOR DE REVOLUȚIE

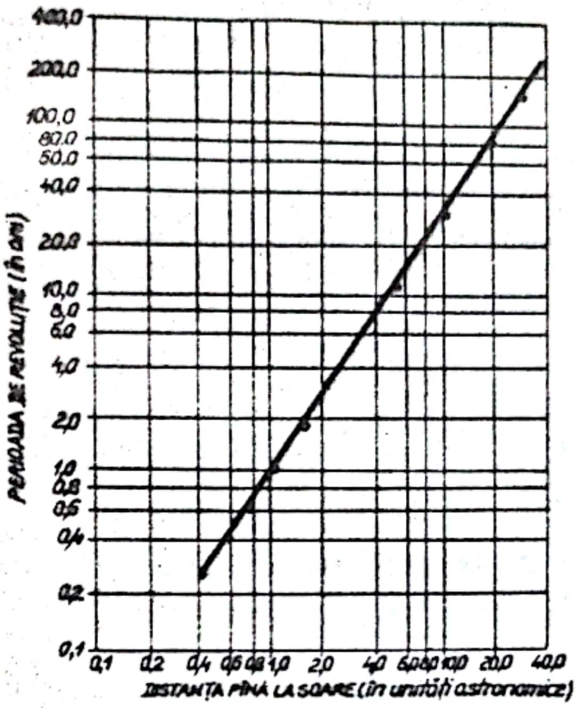


Fig. 15.12. LEGĂTURA DINTRE DISTANȚA FAȚĂ DE SOARE ȘI PERIOADA DE REVOLUȚIE

Din ambele reprezentări grafice se observă că abaterile valorilor empirice față de valorile teoretice sînt foarte mici, ceea ce înseamnă că funcția de tip putere este foarte potrivită pentru a descrie legătura dintre variabile. Datorită acestui fapt în aplicarea metodei punctelor selecționate, pentru estimarea parametrilor acestei ecuații de regresie, se pot utiliza coordonatele celor două puncte empirice extreme (0,39; 0,24) și (39,50 ; 248) în locul unor valori de pe curba trasată vizual pe grafic. Substituind pe rînd aceste valori în ecuația de regresie, rezultă următorul sistem de ecuații normale:

$$0,24 = a \cdot 0,39^b$$

$$248 = a \cdot 39,5^b$$

Împărțind ecuația a doua la prima, rezultă:

$$1033 = 101,3^b$$

de unde, prin logaritmare, se obține:

$$\log 1033 = b \log 101,3$$

iar

$$b = \frac{3,01368}{2,00561} = 1,5$$

Substituind valoarea lui b în una dintre cele două ecuaţii se obţine valoarea parametrului a:

$$0,24 = a.0,39^{1,5}$$

$$\log 0,24 = \log a + 1,5 \log 0,39$$

iar

$$\log a = \log 0,24 - 1,5.\log 0,39 = - 0,61979 - 1,5.(-0,40894) =$$

$$= - 0,00638$$

de unde, prin antilogaritmare, rezultă: $a \simeq 1$

Deci ecuaţia de regresie este

$$y = x^{1,5}$$

Pentru o determinare mai riguroasă a parametrilor ecuaţiei de regresie vom aplica metoda celor mai mică pătrate. Parametrii log a şi b se vor determina cu ajutorul formulelor (15.114) şi (15.115). Elementele de calcul necesare, pentru aplicarea acestor formule, se prezintă în tabelul 15.13.

Tabelul 15.13. Calcule preliminare necesare determinării parametrilor curbei logaritmice

Planeta	x_i	y_i	$\log x_i$	$\log y_i$	$(\log x_i)^2$	$(\log x_i)(\log y_i)$
0	1	2	3	4	5	6
Mercur	0,39	0,24	-0,40894	-0,61979	0,16723	0,25346
Venus	0,72	0,62	-0,14267	-0,20761	0,02035	0,02962
Pământ	1,00	1,00	0,00000	-0,00000	0,00000	0,00000
Marte	1,52	1,88	0,18184	0,27416	0,03307	0,04985
Jupiter	5,20	11,90	0,71600	1,07555	0,51266	0,77009
Saturn	9,54	29,50	0,97955	1,46982	0,95952	1,43976
Uranus	19,20	84,00	1,28330	1,92428	1,64686	2,46943
Neptun	30,10	165,00	1,47857	2,21748	2,18617	3,27870
Pluto	39,50	248,00	1,59660	2,39445	2,54913	3,82298
Total	107,17	542,14	5,66425	8,52834	8,07499	12,11389

$$\log a = \frac{8,52834.8,07499-12,11389-5.68425}{9.8,07499-5,68425^2} = \frac{0,00788}{40,36421} = 0,000195$$

$$b = \frac{9.12,11389-5,68425.8,52834}{9.8,07499-5,68425^2} = \frac{60,54779}{40,36421} = 1,50004 \text{ ani}$$

Ecuaţia de regresie va fi de forma:

$$\log y_x = 0,00019 + 1,50004 (\log x_i)$$

Valorile perioadelor de revoluţie, în funcţie de distanţele faţă de soare, se prezintă în tabelul 15.14.

Tabelul 15.14

Planeta	x_i	y_i	$\log y_{x_i}$	y_{x_i}	$(\log y_i)^2$
Mercur	0,39	0,24	- 0,6198	0,24365	0,38414
Venus	0,72	0,62	- 0,2076	0,61120	0,04310
Pămînt	1,00	1,00	0,0000	1,00044	0,00000
Marte	1,52	1,88	0,2742	1,87483	0,07516
Jupiter	5,20	11,90	1,0755	11,86386	1,15680
Saturn	9,54	29,50	1,4698	29,48181	2,16038
Uranus	19,20	84,00	1,9243	84,17769	3,70285
Neptun	30,10	165,00	2,2175	165,23518	4,91723
Pluto	39,50	248,00	2,3945	248,40091	5,73340
Total	107,17	542,14	8,5283	542,88968	18,17305

Se observă că suma logaritmilor valorilor empirice este egală cu suma logaritmilor valorilor teoretice : $\sum \log y_i = \sum \log y_{x_i}$. Menţionăm că, în cazul regresiei simple după o funcţie putere, nu se verifică relaţia $\sum y_i = \sum y_{x_i}$ din cazul regresiei simple liniare.

Pe baza ecuaţiei de regresie se poate determina care este durata perioadei de revoluţie pentru oricare dintre planetele sistemului solar. De exemplu, dacă ar exista o planetă (I) la distanţa de 60 de unităţi astronomice faţă de soare, aceasta ar efectua o mişcare de revoluţie în jurul soarelui într-un număr de ani dat de ecuaţia:

$$y = 60^{1,5}$$

$$\log y = 1,5 \log 60 = 1,5.1,77815 = 2,66722$$

de unde $y = 465$ ani

Aceasta înseamnă că o planetă care s-ar afla la distanţa de 60 de unităţi astronomice faţă de soare, ar efectua o mişcare de revoluţie în decurs de 465 ani.

Intensitatea legăturii dintre distanţa planetelor faţă de soare şi perioada lor de revoluţie se măsoară cu ajutorul coeficientului de corelaţie, calculat pe baza logaritmilor celor două variabile (15.122) :

$$r_{\log x, \log y} = \frac{9.12,11389 - 5,68425.8,52834}{\sqrt{(9.8,07499-5,68425^2)(9.18,17304-8,52834^2)}} =$$

$$= \frac{60,54779}{60,54808} = 0,9999$$

Aceasta înseamnă că legătura dintre distanţa planetelor faţă de soare şi perioada lor de revoluţie este directă şi foarte strînsă, am putea spune funcţională.

15.1.4. Legături de tip exponenţial

Funcţia exponenţială de tip clasic este de forma:

$$y = a\, e^{bx} \qquad (-\infty < x < \infty) \qquad (15.125)$$

Curba reprezentativă a funcţiei exponenţiale se prezintă în fig. 15.13.

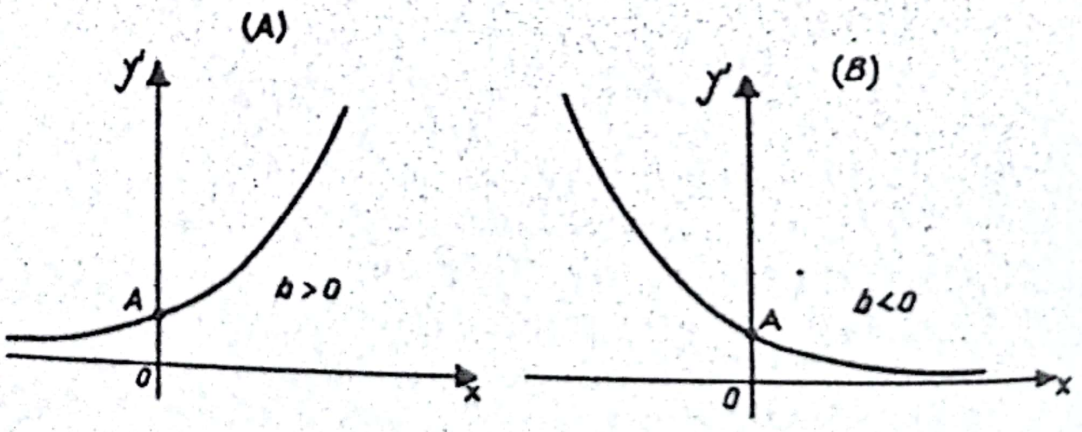


Fig. 15.13. FUNCTIE EXPONENTIALĂ $y = ae^{bx}$

Din fig. 15.13 se observă că toate curbele exponenţiale trec prin punctul de coordonate (o,a), iar axa abciselor este o asimptotă orizontală. Dacă coeficientul de regresie b este pozitiv, atunci curba exponenţială este crescătoare (fig. 15.13 (A)) iar dacă b este negativ, curba exponenţială este descrescătoare (fig. 15.13 (B)).

Liniarizarea legăturii (15.125) se obţine prin substituţia $y_i = \ln y_i$ şi $a = \ln a'$.

Intensitatea legăturii dintre variabile se poate acum determina cu ajutorul coeficientului de corelaţie ce se calculează după una din formulele:

- în cazul perechilor x_i, y_i neponderate:

$$r_{x,\ln y} = \frac{n\sum x_i \ln y_i - \sum x_i \sum \ln y_i}{\sqrt{\left[n\sum x_i^2 - (\sum x_i)^2\right]\left[n\sum(\ln y_i)^2 - (\sum \ln y_i)^2\right]}} \qquad (15.126)$$

- în cazul perechilor de valori corelate x_i, y_i ponderate cu n_i:

$$r_{x,\ln y} = \frac{\sum n_i \sum n_i x_i \ln y_i - \sum x_i n_i \sum (\ln y_i) n_i}{\sqrt{\left[\sum n_i \sum x_i^2 n_i - (\sum x_i n_i)^2\right]\left[\sum n_i (\ln y_i)^2 - (\sum n_i \ln y_i)^2\right]}} \qquad (15.127)$$

- în cazul unui tabel de corelaţie:

$$r_{x,\ln y} = \frac{\sum\sum n_{ij} \sum\sum n_{ij}\, x_i \ln y_j - \sum x_i n_{i.} \sum n_{.j} \ln y_j}{\sqrt{\left[\sum\sum n_{ij} \sum n_{i.} x_i^2 - (\sum n_{i.} x_i)^2\right]\left[\sum\sum n_{ij} \sum n_{.j} (\ln y_j)^2 - (\sum n_{.j} \ln y_j)^2\right]}} \qquad (15.128)$$

O altă formă a fucnţiei exponenţiale, care se situează numai în cadranul întîi al axelor de coordonate, este dat de ecuaţia:

$$y = a\, e^{\frac{b}{x}} \qquad (0 < x < \infty) \qquad (15.129)$$

Curba reprezentativă a funcţiei (15.129) se prezintă în fig. 15.14.

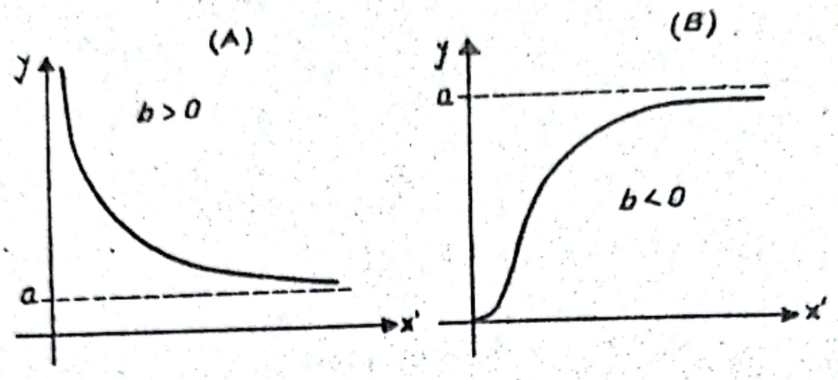


Fig.15.14. FUNCȚIA EXPONENȚIALĂ
$y = ae^{bx}$

Dacă coeficientul de regresie b este pozitiv, curba exponențială este descrescătoare (fig. 15.14(A)), iar dacă b este negativ curba exponențială este crescătoare (fig. 15.14(B)).

Curba exponențială (A) are două asimptote :una orizontală $y = a$ și alta verticală $x = o$.

Curba exponențială (B) se caracterizează prin următoarele trăsături:

- trece prin originea axelor de coordonate;
- are o asimptotă orisontală $y = a$;
- are un punct de inflexiune de coordonate $(-\frac{b}{a}, \frac{a}{e^2})$.

Liniarizarea legăturii expenențiale dintre variabilele (15.125) se obține prin următoarele substituții: $y'_i = \ln y_i$ și $x'_i = \frac{1}{x_i}$ în ecuația $y' = a'e^{bx}$.

Intensitatea legăturii dintre variabile, sub formă liniarizată, se poate calcula cu ajutorul coeficientului de corelație liniară între $\frac{1}{x_i}$ și $\ln y_i$. Corespunzător celor trei modalități de prezentare a datelor statistice,coeficientul de corelație va lua următoarele expresii:

- în primul caz:

$$r_{\frac{1}{x},\ln y} = \frac{n\sum\frac{\ln y_i}{x_i} - \sum\frac{1}{x_i}\sum\ln y_i}{\sqrt{\left[n\sum\frac{1}{x_i^2} - \left(\sum\frac{1}{x_i}\right)^2\right]\left[n\sum(\ln y_i)^2 - \left(\sum\ln y_i\right)^2\right]}} \qquad (15.130)$$

- în al doilea caz:

$$r_{\frac{1}{x},\ln y} = \frac{\sum n_i \sum \frac{n_i \ln y_i}{x_i} - \sum \frac{n_i}{x_i} \sum n_i \ln y_i}{\sqrt{\left[\sum n_i \sum \frac{n_i}{x_i^2} - \left(\sum \frac{n_i}{x_i}\right)^2\right]\left[\sum n_i \sum n_i (\ln y_i)^2 - \left(\sum n_i \ln y_i\right)^2\right]}} \qquad (15.131)$$

- în al treilea caz:

$$r_{\frac{1}{x},\ln y} = \frac{\sum\sum n_{ij} \sum\sum \frac{\ln y_j}{x_i} n_{ij} - \sum \frac{n_{i.}}{x_i} \sum n_{.j} \ln y_j}{\sqrt{\left[\sum\sum n_{ij} \sum \frac{n_{i.}}{x_i^2} - \left(\sum \frac{n_{i.}}{x_i}\right)^2\right]\left[\sum\sum n_{ij} \sum n_{.j} (\ln y_j)^2 - \left(\sum n_{.j} \ln y_j\right)^2\right]}} \qquad (15.132)$$

O altă formă modificată a funcţiei exponenţiale este următoarea:

$$y = \frac{1}{a + b\, e^{-x}} \qquad (-\infty < x < +\infty) \qquad (15.133)$$

Curba exponenţială descrisă de funcţia (15.133) se prezintă în fig. 15.15.

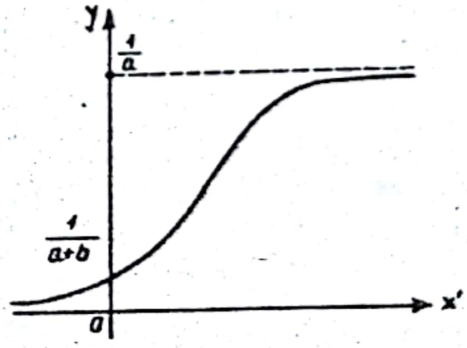


Fig 15.15. CURBA EXPONENŢIALĂ $y = \frac{1}{a+b^{-x}}$

Se observă, din fig. 15.15, că această curbă exponenţială are două asimptote. orizontale una este chiar axa abscielor y=0 şi cealaltă, paralelă cu cea dintîi , este $y = \frac{1}{a}$. Curba exponenţială intersectează axa coordonatelor în punctul de coordonate

$(0, \frac{1}{a+b})$ şi are un punct de inflexiune $(\ln\frac{b}{a}, \frac{1}{2a})$.

Mai des utilizate în practică sînt două forme particulare ale acestei curbe şi anume curba lui Gompertz şi curba logistică de care ne vom ocupa în capitolul regeritor la analiza seriilor dinamice.

Liniarizarea legăturii exponenţiale are loc dacă, în formula (15.133) se fac subsituţiile: $y = \frac{1}{y_1}$ şi are $x' = e^{-x}$.

Intensitatea legăturii, în acest tip de funcţie exponenţială, după liniarizare, se poate determina cu ajutorul uneia din următoarele formule, în baza modului de prezentare a datelor empirice:

- cînd perechile de valori x_i, y_i sînt neponderate:

$$r_{e^{-x},\frac{1}{y}} = \frac{n\sum(e^{-x_i})\frac{1}{y_i} - \sum e^{-x_i}\sum\frac{1}{y_i}}{\sqrt{\left[n\sum(e^{-x_i})^2 - (\sum e^{-x_i})^2\right]\left[n\sum\frac{1}{y_i^2} - (\sum\frac{1}{y_i})^2\right]}} \qquad (15.134)$$

- cînd perechile de valori x_i, y_i sînt ponderate cu n_i:

$$r_{e^{-x},\frac{1}{y}} = \frac{\sum n_i\sum n_i(e^{-x_i})\frac{1}{y_i} - \sum n_i e^{-x_i}\sum\frac{1}{y_i}}{\sqrt{\left[\sum n_i\sum n_i(e^{-x_i})^2 - (\sum n_i e^{-x_i})^2\right]\left[\sum n_i\sum\frac{n_i}{y_i^2} - (\sum\frac{n_i}{y_i})^2\right]}} \qquad (15.135)$$

- cînd variabilele X şi Y se prezintă sub forma unui tabel de corelaţie:

$$r_{e^{-x},\frac{1}{y}} = \frac{\sum\sum n_{ij}\sum\sum n_{ij}(e^{-x_i})\frac{1}{y_i} - \sum n_{i.}e^{-x_i}\sum\frac{n_{.i}}{y_i}}{\sqrt{\left[\sum\sum n_{ij}\sum n_{i.}(e^{-x_i})^2 - (\sum n_{i.}e^{-x_i})^2\right]\left[\sum\sum n_{ij}\sum\frac{n_{.i}}{y_j^2} - (\sum\frac{n_{.i}}{y_j})^2\right]}} \qquad (15.136)$$

O formă transformată a ecuaţiei exponenţiale, convenabilă pentru aplicaţiile practice, este:

sau $$y = a\,b^x \qquad (15.137)$$

$$\log y = \log a + x \log b$$

Pentru determinarea parametrilor ecuaţiei de regresie (15.137), prin aplicarea metodei celor mai mici pătrate, se obţine sistemul de ecuaţii normale:

$$\begin{aligned} n \log a + \log b \sum x_i &= \sum \log y_i \\ \log a \sum x_i + \log b \sum x_i^2 &= \sum x_i \log y_i \end{aligned} \qquad (15.138)$$

Prin rezolvarea sistemului (15.138) se obţin formulele uzuale ale parametrilor ecuaţiei de regresie.

Sistemul (15.138) este aplicabil în primul caz de sistematizare a datelor cu privire la cele două variabile. In celelalte două cazuri, sistemul ia forma:

- în cazul al doilea:

$$\begin{aligned} \log a \sum n_i + \log b \sum x_i^2 n_i &= \sum x_i \log y_i \\ \log a \sum x_i n_i + \log b \sum x_i^2 n_i &= \sum n_i x_i \log y_i \end{aligned} \qquad (15.139)$$

- în cazul al treilea:

$$\begin{aligned} \log a \sum\sum n_{ij} + \log b \sum x_i n_{i.} &= \sum n_{.j} \log y_j \\ \log a \sum x_i n_{i.} + \log b \sum x_i^2 n_{i.} &= \sum\sum n_{ij} x_i \log y_j \end{aligned} \qquad (15.140)$$

Intensitatea corelaţiei exponenţiale (15.137) se determină cu ajutorul raportului de corelaţie. Raportul de corelaţie, în funcţie de modul de prezentare a datelor iniţiale, se determină astfel:

- în primul caz:

$$\eta = \sqrt{\frac{a\sum \log y_i + \log b \sum x_i \log y_i - \frac{(\sum \log y_i)^2}{n}}{\sum (\log y_i)^2 - \frac{(\sum \log y_i)^2}{n}}} \qquad (15.141)$$

- în cazul al doilea:

$$\eta = \sqrt{\frac{a\sum n_i \log y_i + \log b \sum n_i x_i \log y_i - \frac{(\sum n_i \log y_i)^2}{\sum n_i}}{\sum n_i (\log y_i)^2 - \frac{(\sum n_i \log y_i)^2}{\sum n_i}}} \qquad (15.142)$$

- în cazul al treilea:

$$\eta = \sqrt{\frac{a\sum n_{\cdot j}\log y_j + \log b \sum\sum n_{ij}x_i \log y_j - \frac{(\sum n_{\cdot j}\log y_j)^2}{\sum\sum n_{ij}}}{\sum n_{\cdot j}(\log y_j)^2 - \frac{(\sum n_{\cdot j}\log y_j^2)}{\sum\sum n_{ij}}}} \qquad (15.143)$$

Dacă în ecuaţia (15.137) se face substituţia $\log y' = y$, se liniarizează legătura exponenţială. Intensitatea legăturii se poate măsura cu ajutorul coeficientului de corelaţie liniară dintre x şi $\log y$. Corespunzător celor trei cazuri de sistematizare a datelor iniţiale, formula coeficientului de corelaţie devine:

- în primul caz:

$$r_{x,\log y} = \frac{n\sum x_i \log y_i - \sum x_i \sum \log y_i}{\sqrt{\left[n\sum x_i^2 - (\sum x_i)^2\right]\left[n\sum y_i^2 - (\sum y_i)^2\right]}} \qquad (15.144)$$

- în cazul al doilea:

$$r_{x,\log y} = \frac{\sum n_i \sum n_i x_i \log y_i - \sum n_i x_i \sum n_i \log y_i}{\sqrt{\left[\sum n_i \sum n_i x_i^2 - (\sum n_i x_i)^2\right]\left[\sum n_i \sum n_i (\log y_i)^2 - (\sum n_i (\log y_i)^2\right]}} \qquad (15.145)$$

- în cazul al treilea:

$$\log y = \frac{\sum\sum n_{ij} \sum\sum n_{ij} x_i \log y_i - \sum n_{i\cdot} x_i \sum n_{\cdot j} \log y_j}{\sqrt{\left[\sum\sum n_{ij} \sum x_i^2 n_{i\cdot} - (\sum x_n n_{i\cdot})^2\right]\left[\sum\sum n_{ij} \sum n_{\cdot j} (\log y_j)^2 - (\sum n_{\cdot j} \log y_j)^2\right]}} \qquad (15.146)$$

15.1.5. Legături de tip logaritmic

Funcţia logaritmică este dată de expresia:

$$y = a + b \log x \qquad (0 < x < \infty) \qquad (15.147)$$

- în cazul al treilea:

$$\eta = \sqrt{\frac{a\sum n_{\cdot j}\log y_j + \log b \sum\sum n_{ij}x_i \log y_j - \frac{(\sum n_{\cdot j}\log y_j)^2}{\sum\sum n_{ij}}}{\sum n_{\cdot j}(\log y_j)^2 - \frac{(\sum n_j \log y_j^2)}{\sum\sum n_{ij}}}} \qquad (15.143)$$

Dacă în ecuaţia (15.137) se face substituţia $\log y' = y$, se liniarizează legătura exponenţială. Intensitatea legăturii se poate măsura cu ajutorul coeficientului de corelaţie liniară dintre x şi $\log y$. Corespunzător celor trei cazuri de sistematizare a datelor iniţiale, formula coeficientului de corelaţie devine:

- în primul caz:

$$r_{x,\log y} = \frac{n\sum x_i \log y_i - \sum x_i \sum \log y_i}{\sqrt{\left[n\sum x_i^2 - (\sum x_i)^2\right]\left[n\sum y_i^2 - (\sum y_i)^2\right]}} \qquad (15.144)$$

- în cazul al doilea:

$$r_{,\log y} = \frac{\sum n_i \sum n_i x_i \log y_i - \sum n_i x_i \sum n_i \log y_i}{\sqrt{\left[\sum n_i \sum n_i x_i^2 - (\sum n_i x_i)^2\right]\left[\sum n_i \sum n_i (\log y_i)^2 - (\sum n_i (\log y_i)^2\right]}} \qquad (15.145)$$

- în cazul al treilea:

$$\log y = \frac{\sum\sum n_{ij} \sum\sum n_{ij} x_i \log y_i - \sum n_{i\cdot} x_i \sum n_{\cdot j} \log y_i}{\sqrt{\left[\sum\sum n_{ij} \sum x_i^2 n_{i\cdot} - (\sum x_n n_{i\cdot})^2\right]\left[\sum\sum n_{ij} \sum n_{\cdot j} (\log y_j)^2 - (\sum n_{\cdot j} \log y_j)^2\right]}} \qquad (15.146)$$

15.1.5. Legături de tip logaritmic

Funcţia logaritmică este dată de expresia:

$$y = a + b \log x \qquad (0 < x < \infty) \qquad (15.147)$$

- în cazul al treilea:

$$\eta = \sqrt{\frac{a\sum n_{\cdot j}\log y_j + \log b \sum\sum n_{ij}x_i \log y_j - \frac{(\sum n_{\cdot j}\log y_j)^2}{\sum\sum n_{ij}}}{\sum n_{\cdot j}(\log y_j)^2 - \frac{(\sum n_j \log y_j^2)}{\sum\sum n_{ij}}}} \qquad (15.143)$$

Dacă în ecuaţia (15.137) se face substituţia $\log y' = y$, se liniarizează legătura exponenţială. Intensitatea legăturii se poate măsura cu ajutorul coeficientului de corelaţie liniară între x şi $\log y$. Corespunzător celor trei cazuri de sistematizare a datelor iniţiale, formula coeficientului de corelaţie devine:

- în primul caz:

$$r_{x,\log y} = \frac{n\sum x_i \log y_i - \sum x_i \sum \log y_i}{\sqrt{\left[n\sum x_i^2 - (\sum x_i)^2\right]\left[n\sum y_i^2 - (\sum y_i)^2\right]}} \qquad (15.144)$$

- în cazul al doilea:

$$r_{x,\log y} = \frac{\sum n_i \sum n_i x_i \log y_i - \sum n_i x_i \sum n_i \log y_i}{\sqrt{\left[\sum n_i \sum n_i x_i^2 - (\sum n_i x_i)^2\right]\left[\sum n_i \sum n_i (\log y_i)^2 - (\sum n_i (\log y_i)^2\right]}} \qquad (15.145)$$

- în cazul al treilea:

$$\log y = \frac{\sum\sum n_{ij} \sum\sum n_{ij} x_i \log y_i - \sum n_{i\cdot} x_i \sum n_{\cdot j} \log y_j}{\sqrt{\left[\sum\sum n_{ij} \sum x_i^2 n_{i\cdot} - (\sum x_n n_{i\cdot})^2\right]\left[\sum\sum n_{ij} \sum n_{\cdot j} (\log y_j)^2 - (\sum n_{\cdot j} \log y_j)^2\right]}} \qquad (15.146)$$

15.1.5. Legături de tip logaritmic

Funcţia logaritmică este dată de expresia:

$$y = a + b \log x \qquad (0 < x < \infty) \qquad (15.147)$$

Curba funcţiei (15.147) se prezintă în fig. 15.16.

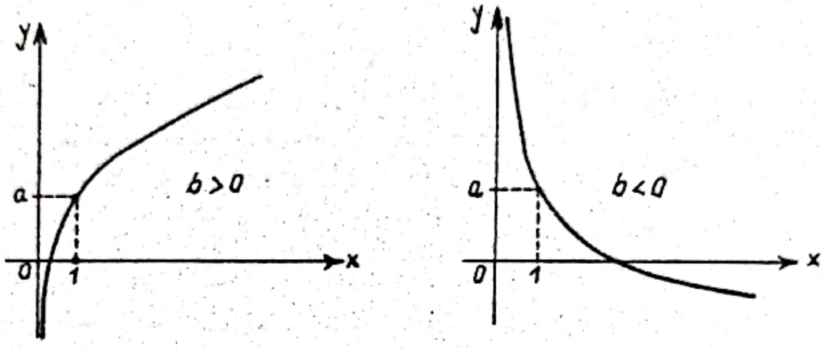


Fig. 15.16 CURBA LOGARITMICĂ $y = a + b \log x$

Din fig. 15.16 se observă că, dacă coeficientul de regresie b este pozitiv, curba este crescătoare (fig. 15.16 (A)), iar dacă b este negativ, curba este descrescătoare (fig. 15.16 (B)). De asemenea, ambele curbe trec prin punctul de coordonate (1,a) şi au ca asimptotă verticală axa coordonatelor $x = 0$.

Parametrii funcţiei de regresie (15.147) se pot determina cu ajutorul metodei celor mai mici pătrate. Prin aplicarea acestei metode se obţin, în funcţie de modul de prezentare a celor două variabile corelate, următoarele sisteme de ecuaţii normale:

- cînd perechile de valori x_i, y_i sînt neponderate:

$$\begin{aligned} n\,a + b\sum \log x_i &= \sum y_i \\ a\sum \log x_i + b\sum (\log x_i)^2 &= \sum y_i \log x_i \end{aligned} \qquad (15.148)$$

iar

$$a = \frac{\sum y_i \sum (\log x_i)^2 - \sum \log x_i \sum y_i \log x_i}{n\sum (\log x_i)^2 - (\sum \log x_i)^2} \qquad (15.149)$$

$$b = \frac{n\sum y_i \log x_i - \sum \log x_i \sum y_i}{n\sum (\log x_i)^2 - (\sum \log x_i)^2} \qquad (15.150)$$

- cînd perechile de valori corelate x_i, y_i sînt ponderate cu frecvenţa n_i:

$$a\sum n_i + b\sum n_i \log x_i = \sum n_i y_i \qquad (15.151)$$
$$a\sum n_i \log x_i + b\sum n_i (\log x_i)^2 = \sum n_i y_i \log x_i$$

de unde rezultă:

$$a = \frac{\sum n_i y_i \sum n_i (\log x_i)^2 - \sum n_i \log x_i \sum n_i y_i \log x_i}{\sum n_i \sum n_i (\log x_i)^2 - (\sum n_i \log x_i)^2} \qquad (15.152)$$

$$b = \frac{\sum n_i \sum n_i y_i \log x_i - \sum n_i \log x_i \sum n_i y_i}{\sum n_i \sum n_i (\log x_i)^2 - (\sum n_i \log x_i)^2} \qquad (15.153)$$

- cînd cele două variabile corelate se prezintă sub forma unui tabel cu dublă intrare:

$$a\sum\sum n_{ij} + b\sum n_{i\cdot} \log x_i = \sum n_{\cdot j} y_j \qquad (15.154)$$
$$a\sum n_{i\cdot} \log x_i + b\sum n_{i\cdot} (\log x_i)^2 = \sum\sum n_{ij} y_j \log x_i$$

iar

$$a = \frac{\sum n_{\cdot j} y_j \sum n_{i\cdot} (\log x_i)^2 - \sum n_{i\cdot} \log x_i \sum\sum n_{ij} y_j \log x_i}{\sum\sum n_{ij} \sum n_{i\cdot} (\log x_i)^2 - (\sum n_{i\cdot} \log x_i)^2} \qquad (15.155)$$

$$b = \frac{\sum\sum n_{ij} \sum\sum n_{ij} y_j \log x_i - \sum n_i \log x_i \sum n_{\cdot j} y_j}{\sum\sum n_{ij} \sum n_{i\cdot} (\log x_i)^2 - (\sum n_{i\cdot} \log x_i)^2} \qquad (15.156)$$

Intensitatea legăturii logaritmice de tipul (15.148) se măsoară cu ajutorul raportului de corelaţie:

$$\eta = \sqrt{1 - \frac{\sigma^2_{yy_x}}{\sigma^2_y}} \quad \text{în care} \quad y_x = a + b \log x \qquad (15.157)$$

Calculul simplificat al raportului de corelaţie în funcţie de forma de prezentare a celor două variabile corelate, se efectuează după una din următoarele formule:

- în primul caz:

$$\eta = \sqrt{\frac{a\sum y_i + b\sum y_i \log x_i - \frac{(\sum y_i)^2}{n}}{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}} \qquad (15.158)$$

- în al doilea caz:

$$\eta = \sqrt{\frac{a\sum n_i y_i + b\sum n_i y_i \log x_i - \frac{(\sum n_i y_i)^2}{\sum n_i}}{\sum n_i y_i^2 - \frac{(\sum n_i y_i)^2}{\sum n_i}}} \qquad (15.159)$$

- în al treilea caz:

$$\eta = \sqrt{\frac{a\sum n_{\cdot j} y_j + b\sum\sum n_{ij} y_j \log x_i - \frac{(\sum n_{\cdot j} y_i)^2}{\sum\sum n_{ij}}}{\sum\sum n_{ij} \sum n_{\cdot j} y_j^2 - \frac{(\sum n_{\cdot j} y_i)^2}{\sum\sum n_{ij}}}} \qquad (15.160)$$

Ecuaţia logaritmică (15.147) se poate liniariza prin substituţiile :

$$x = \log x' \quad \text{şi} \quad y = y'$$

$$y' = a + b \log x' \longrightarrow y = a + b\,x$$

Coeficientul de corelaţie dintre log x şi y va avea următoarele expresii:

- în cazul x_i, y_i nepondorate

$$r_{\log x, y} = \frac{n\sum y_i \log x_i - \sum \log x_i \sum y_i}{\sqrt{\left[n\sum(\log x_i)^2 - (\sum \log x_i)^2\right]\left[n\sum(\log y_i)^2 - (\sum \log y_i)^2\right]}} \qquad (15.161)$$

în cazul al doilea:

$$r_{\log x, y} = \frac{\sum n_i \sum n_i y_i \log x_i - \sum n_i \log x_i \sum n_i y_i}{\sqrt{\left[\sum n_i \sum n_i (\log x_i)^2 - (\sum n_i \log x_i)^2\right]\left[\sum n_i \sum n_i y_i^2 - (\sum n_i y_i)^2\right]}} \qquad (15.162)$$

- în cazul al treilea:

$$r_{\log x, y} = \frac{\sum\sum n_{ij} \sum\sum n_{ij} y_i \log x_i - \sum n_{i\cdot} \log x_i \sum n_{\cdot j} y_i}{\sqrt{\left[\sum\sum n_{ij} \sum n_{i\cdot} x_i^2 - (\sum n_{i\cdot} x_i)^2\right]\left[\sum\sum n_{ij} \sum n_{\cdot j} y_j^2 - (\sum n_{\cdot j} y_j)^2\right]}} \qquad (15.163)$$

15.2. Alegerea celei mai potrivite ecuaţii de regresie

Una dintre problemele fundamentale ale regresiei şi corelaţiei este alegerea celui mai potrivit model de regresie. De modul cum este rezolvată această problemă depinde încrederea ce se poate acorda indicatorilor obţinuţi din calculele de regresie şi corelaţie.

Dacă există legităţi, pe care le cunoaştem apriori, cu privire la forma legăturii dintre variabile, atunci vom utiliza drept ecuaţie de regresie expresia matematică a acestora. Dacă nu există o asemenea bază teoretică, se procedează la o analiză calitativă a esenţei fenomenelor şi dacă se ajunge la concluzia că există legătură între variabile, pentru a alege cel mai potrivit model de regresie, se pot utiliza următoarele reguli:

1. Se reprezintă grafic datele empirice şi se trasează vizual o curbă care să se apropie cît mai mult de punctele empirice şi apoi se alege o ecuaţie care să corespundă cît mai bine curbei empirice.

Această metodă se poate utiliza în practică numai dacă numărul perechilor de valori este suficient de mare ($n > 30$).

2. Pentru alegerea în fiecare caz particular a curbei celei mai potrivite, se poate utiliza şi metoda diferenţelor succesive ale variabilei Y.

3. Cînd primele diferenţe ale variabilei X sînt egale:

- dacă primele valori ale variabilei Y sînt constante, atunci se recomandă utilizarea modelului liniar de regresie simplă $y = a + bx$;

- dacă diferenţele de ordinul al doilea sînt constante, atunci se recomandă utilizarea modelului parabolei de gradul doi $y = a + b x + c x^2$;

- dacă diferenţele de ordinul al treilea sînt constante, atunci se recomandă utilizarea parabolei de gradul trei $y = a + b x + c x^2 + d x^3$;

- în general, dacă diferenţele de ordinul n sînt constante, atunci se recomandă utilizarea polinomului de gradul n ;

- dacă primele diferenţe ale logaritmilor valorilor lui Y sînt constante, atunci se recomandă utilizarea unei curbe semilogaritmice de forma $\log y = a + bx$;

- dacă primele diferenţe ale mărimilor inverse ale variabilei dependente Y sînt constante, se poate utiliza o curbă hiperbolică de forma : $\frac{1}{y} = a + bx$;

- dacă raportul dintre diferenţele de ordinul întîi ale variabilei X şi cele corespunzătoare pentru variabila dependentă Y este constant, se poate utiliza modelul unei curbe dublu logaritmice de forma:

$$\log y = a + b \log x$$

4. Se reprezintă datele empirice se o hîrtie milimetrică obişnuită:

- dacă corelograma variabilelor X şi log Y ia o formă liniară, atunci legătura dintre variabile se poate aproxima printr-o curbă semilogaritmică:

$$\log y = a + bx$$

- dacă corelograma logaritmilor variabilelor X şi Y ia o formă liniară, atunci legătura dintre variabile poate fi descrisă de o ecuaţie logaritmică

- dacă graficul perechilor de valori x şi $\frac{1}{y}$ ia forma unei drepte atunci cînd legătura dintre variabile este hiperbolică $\frac{1}{y} = a + bx$.

5. Dacă se reprezintă pe o hîrtie semilogaritmică perechile de valori corelate x_1, y_1 şi corelograma ia forma unei drepte, atunci legătura dintre variabile se apropie de o curbă semilogaritmică $\log y = a + bx$

6. Dacă corelograma valorilor empirice ia forma unei drepte pe o reţea dublu logaritmică, atunci legătura dintre variabile ia forma unei funcţii putere: $\log y = \log a + b \log x$

7. Dacă pe un grafic, în care pe abscisă scara este aritmetică iar pe ordonată valorile reciproce, corelaograma ia forma unei drepte, legătura dintre variabile se apropie de o hiperbolă descrisă de ecuaţia $\frac{1}{y} = a + bx$.

Desigur, în activitatea statistică diferenţele de diferite ordine se vor apropia mai mult sau mai puţin de egalitatea între ele şi numai în cazuri cu totul particulare ele vor fi chiar egale. Curbele empirice se vor apropia mai mult sau mai puţin de anumite modele teoretice şi acestea, chiar dacă nu ne vor spune dintr-o dată care este curba cea mai potrivită în fiecare caz particular în parte, ne va ajuta totuşi să excludem din calcul anumite familii de curbe şi apoi, prin încercări succesive, se va preciza care curbă este mai potrivită, cu ajutorul diferitelor tehnici statistice.

Scanned with OKEN Scanner

16. REGRESIE SI CORELATIE MULTIPLA

16.1. Noţiunea de legătură multiplă

Legăturile multiple de corelaţie se stabilesc între două sau mai multe caracteristici factoriale şi o caracteristică rezultativă. Astfel de legături se întîlnesc atît în cadrul fenomenelor naturale cît şi în cadrul fenomenelor sociale şi economice complexe.

Un exemplu clasic de regresie şi corelaţie multiplă este cel care se referă la legătura dintre greutatea unei persoane de un anumit sex (Y), înălţimea (X_1) şi vîrsta (X_2).

Un alt exemplu, corelaţia dintre nivelul de industrializare a zonelor unei ţări (Y) - exprimate prin valoarea producţiei globale industriale pe zone, pe de o parte şi nivelul de înzestrare tehnică a întreprinderilor industriale (X_1) - exprimată prin valoarea fondurilor fixe şi forţa de muncă din industrie (X_2) - exprimată prin numărul personalului angajat în industrie, pe de altă parte.

Modelul acestei legături este:

$$Y = f(X_1, X_2) \qquad (16.1)$$

In agricultură se poate studia corelaţia dintre mărimea venitului net (Y) al C.A.P.- pe de o parte - şi cheltuielile materiale (X_1), producţia globală pe o normă convenţională (X_2), valoarea fondurilor fixe (X_3), după modelul :

$$Y = f(X_1, X_2, X_3) \qquad (16.2)$$

Procesul de urbanizare într-o ţară - exprimat prin ritmul de creştere al populaţiei urbane (Y) - se poate studia în funcţie de următorii factori de influenţă: numărul populaţiei oraşelor în momentul efectuării calculelor (X_1), numărul şi structura forţei de muncă prezentă (X_2), ritmul planificat de dezvoltare a

activităţilor neagricole în oraşe (X_3), distanţa faţă de cele mai apropiate oraşe (X_4) etc.

Creşterea economică se poate exprima printr-o funcţie de regresie după modelul funcţiilor de producţie de tip Cobb-Douglas[1]

$$Y = a X_1^{b_1} X_2^{b_2} \qquad (16.3)$$

în care variabilele reprezintă următorii indicatori:

Y = venitul naţional;

X_1 = valoarea fondurilor fixe productive din cadrul economiei naţionale;

X_2 = forţa de muncă din sfera productivă.

Parametrii b_1 şi b_2 pot să reprezinte elasticităţi iar suma lor este egală cu unitatea dacă se consideră că Y depinde numai de X_1 şi X_2: $b_1+b_2 = 1$ de unde $b_1 = 1-b_2$ şi $b_2 = 1-b_1$.

Ul alt economist, Edelberg (1936) a introdus un al treilea factor - durata procesului de producţie X_3:

$$Y = aX_1^{b_1} X_2^{b_2} X_3^{b_3} \qquad (16.4)$$

în care $b_1+b_2+b_3 = 1$.

Cercetările ulterioare au arătat că acest tip de funcţii se poate aplica atît la nivel microeconomic, cît şi la nivel macro.

Remarcăm şi faptul că pentru utilizarea în practică a acestor funcţii de regresie se poate proceda la liniarizarea lor prin logaritmare.

In cadrul legăturilor multiple variabilele factoriale au influenţe diferite asupra variabilei rezultative, unele exercită o acţiune importantă asupra fenomenului efect şi trebuie să fie luate în calculele de regresie şi corelaţie, în timp ce altele au o acţiune mai puţin importantă şi pot fi neglijate. Modelele de corelaţie, ca oricare alte modele simplifică realitatea, deoarece este foarte dificil să se ia în consideraţie mulţimea tuturor factorilor cauzali care acţionează asupra unui anumit fenomen. De aceea, factorii cauzali se ierarhizează în ordinea importanţei acţiunii lor asupra fenomeneului efect şi se iau în

1) Cobb-Douglas: A Theory of Production, în "American Economic Review", nr. 18, 1928, pp. 139-165.

calcul numai cei care au o acţiune importantă.

Pentru ilustrarea calculului şi analizei statistice a corelaţiei multiple dintre fenomene, să considerăm o variabilă dependentă Y şi k variabile independente $(X_1, X_2, \ldots, X_i, \ldots, X_k)$ legate între ele printr-o funcţie:

$$Y = f(X_i) \quad i=1,2,\ldots,k \quad j=1,2,\ldots,n \qquad (16.5)$$

Fiecare dintre variabilele X_i va lua n valori. Ecuaţia (16.5) poate fi liniară sau curbilinie în funcţie de formele legăturilor dintre fiecare pereche de valori (Y, X_i). Dacă toate legăturile simple dintre Y şi X_i sînt liniare, atunci şi legătura multiplă de corelaţie este tot liniară, iar dacă cel puţin una dintre legăturile simple (Y, X_i) este curbilinie, atunci şi legătura multiplă este curbilinie.

16.2. Regresie şi corelaţie liniară multiplă

16.2.1. Regresie liniară multiplă

In cazul unei legături liniare multiple, ecuaţia (16.5) reprezintă o funcţie liniară de forma:

$$Y_{x_1,x_2,\ldots,x_k} = a_0 + a_1X_1 + a_2X_2 + \ldots + a_iX_i + \ldots + a_kX_k \qquad (16.6)$$

în care

a_0 - reprezintă parametrul care exprimă influenţa celorlalţi factori consideraţi cu acţiune constantă, în afară de cei n factori cauzali luaţi în calcul;

a_i, $i=1,2,\ldots,k$ - reprezintă coeficienţii de regresie multiplă care arată cu cît variază variabila rezultativă, atunci cînd variabila factorială X_i se modifică cu o unitate.

Legătura multiplă liniară se reprezintă grafic sub forma unui plan. In cazul corelaţiei dintre două variabile independente (X_1 şi X_2) şi o variabilă rezultativă Y, expresia geometrică a legăturii se prezintă sub forma planului ABCD din fig. 16.1.

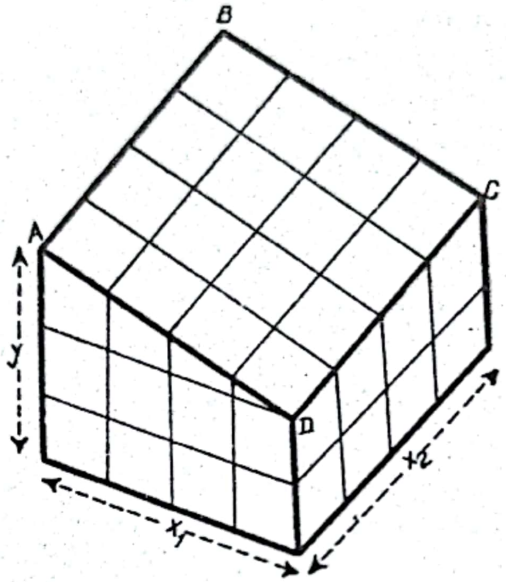


Fig.16.1. CAZUL TIPIC AL UNEI CORELAȚII MULTIPLE LINIARE

Din modelul (16.6) se observă că în cazul a k variabile factoriale ecuația de regresie multiplă liniară depinde de k+1 parametri: $a_0, a_1, a_2, \ldots, a_i, \ldots, a_k$.

Parametrii $a_0, a_1, a_2, \ldots, a_i, \ldots, a_k$ se determină cu ajutorul metodei celor mai mici pătrate, punînd condiția ca suma pătratelor abaterilor dintre valorile empirice y_i și valorile teoretice corespunzătoare $y_{x_1, x_2, \ldots, x_k}$ să fie minimă:

$$S = \sum_j (y_j - y_{x_1, x_2, \ldots, x_k})^2 = \text{minim} \qquad j=1,2,\ldots,n \qquad (16.7)$$

sau

$$S = \sum_j (y_j - a_0 - a_1 X_1 - a_2 X_2 - \ldots - a_i X_i - \ldots - a_k X_k)^2 = \text{minim}$$

Pentru a afla minimul funcției S se determină derivatele parțiale ale acestei funcții în raport cu parametrii $a_0, a_1, \ldots, a_i, \ldots, a_k$ care apoi se egalează cu zero și se obține:

$$\frac{\partial S}{\partial a_0} = 2\sum(Y_j - a_0 - a_1X_1 - a_2X_2 - \ldots - a_iX_i - \ldots - a_kX_k)(-1) = 0$$

$$\frac{\partial S}{\partial a_1} = 2\sum(Y_j - a_0 - a_1X_1 - a_2X_2 - \ldots - a_iX_i - \ldots - a_kX_k)(-X_1) = 0$$

$$\frac{\partial S}{\partial a_2} = 2\sum(Y_j - a_0 - a_1X_1 - a_2X_2 - \ldots - a_iX_i - \ldots - a_kX_k)(-X_2) = 0 \qquad (16.8)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\frac{\partial S}{\partial a_i} = 2\sum(Y_j - a_0 - a_1X_1 - \ldots - a_iX_i - \ldots - a_kX_k)(-X_i) = 0$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\frac{\partial S}{\partial a_k} = 2\sum(Y_j - a_0 - a_1X_1 - a_2X_2 - \ldots - a_iX_i - \ldots - a_kX_k)(-X_k) = 0$$

Prin transformări elementare, sistemul (16.8) devine:

$$na_0 + a_1\sum X_1 + a_2\sum X_2 + \ldots + a_i\sum X_i + \ldots + a_k\sum X_k = \sum Y$$

$$a_0\sum X_1 + a_1\sum X_1^2 + a_2\sum X_2X_1 + \ldots + a_i\sum X_iX_1 + \ldots + a_k\sum X_kX_1 = \sum YX_1$$

$$a_0\sum X_2 + a_1\sum X_1X_2 + a_2\sum X_2^2 + \ldots + a_i\sum X_iX_2 + \ldots + a_k\sum X_kX_2 = \sum YX_2 \qquad (16.9)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$a_0\sum X_i + a_1\sum X_1X_i + a_2\sum X_2X_i + \ldots + a_i\sum X_i^2 + \ldots + a_k\sum X_kX_i = \sum YX_i$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$a_0\sum X_k + a_1\sum X_1X_k + a_2\sum X_2X_k + \ldots + a_i\sum X_iX_k + \ldots + a_k\sum X_k^2 = \sum YX_k$$

Dacă se împarte prima ecuaţie a sistemului (16.9) la n rezultă:

$$a_0 + a_1\bar{x}_1 + a_2\bar{x}_2 + \ldots + a_i\bar{x}_i + \ldots + a_k\bar{x}_k = \bar{y} \qquad (16.10)$$

Aceasta înseamnă că planul de regresie trece prin punctul de coordonate $(\bar{y}, \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_i, \ldots, \bar{x}_k)$.

In cazul a două variabile independente şi una dependentă, sistemul de ecuaţii normale (16.9) se poate scrie substituind fiecare variabilă prin abaterile acesteia faţă de medie. Astfel, în loc de $\sum Y$ se poate scrie $\sum(Y-\bar{y})$, în loc de $\sum X_1$ se poate scrie $\sum(X_1-\bar{x}_1)$, iar în loc de $\sum X_2$ se poate scrie $\sum(X_2-\bar{x}_2)$ şi aşa mai departe.

Această transformare prezintă avantajul simplificării calculelor, deoarece se ştie că suma abaterilor unei variabile faţă de media lor este egală cu zero, adică: $\sum(X_2-\bar{x}_2) = \sum(Y-\bar{y}) = 0$.

In aceste condiţii, prima ecuaţie din sistem se elimină şi sistemul se poate scrie:

$$\begin{aligned} a_1 \sum(X_1-\bar{x}_1)^2 + a_2 \sum(X_1-\bar{x}_1)(X_2-\bar{x}_2) &= \sum(X_1-\bar{x}_1)(Y-\bar{y}_1) \\ a_1 \sum(X_1-\bar{x}_1)(X_2-\bar{x}_2) + a_2 \sum(X_2-\bar{x}_2)^2 &= \sum(X_2-\bar{x}_2)(Y-\bar{y}) \end{aligned} \qquad (16.11)$$

Rezolvînd sistemul (16.11) prin metoda determinanţilor rezultă:

$$a_1 = \frac{\begin{vmatrix} \sum(X_1-\bar{x}_1)(Y-\bar{y}) & \sum(X_1-\bar{x}_1)(X_2-\bar{x}_2) \\ \sum(X_2-\bar{x}_2)(Y-\bar{y}) & \sum(X_2-\bar{x}_2)^2 \end{vmatrix}}{\begin{vmatrix} \sum(X_1-\bar{x}_1)^2 & \sum(X_1-\bar{x}_1)(X_2-\bar{x}_2) \\ \sum(X_1-\bar{x}_1)(X_2-\bar{x}_2) & \sum(X_2-\bar{x}_2)^2 \end{vmatrix}} =$$

$$= \frac{\sum(X_1-\bar{x}_1)(Y-\bar{y})\sum(X_2-\bar{x}_2)^2 - \sum(X_2-\bar{x}_2)(Y-\bar{y})\sum(X_1-\bar{x}_1)(X_2-\bar{x}_2)}{\sum(X_1-\bar{x}_1)^2 \sum(X_2-\bar{x}_2)^2 - \left[\sum(X_1-\bar{x}_1)(X_2-\bar{x}_2)\right]^2} \qquad (16.12)$$

şi

$$a_2 = \frac{\begin{vmatrix} \sum(X_1-\bar{x}_1)^2 & \sum(X_1-\bar{x}_1)(Y-\bar{y}) \\ \sum(X_1-\bar{x}_1)(X_2-\bar{x}_2) & \sum(X_2-\bar{x}_2)(Y-\bar{y}) \end{vmatrix}}{\begin{vmatrix} \sum(X_1-\bar{x}_1)^2 & \sum(X_1-\bar{x}_1)(X_2-\bar{x}_2) \\ \sum(X_1-\bar{x}_1)(X_2-\bar{x}_2) & \sum(X_2-\bar{x}_2)^2 \end{vmatrix}} =$$

$$= \frac{\sum(X_1-\bar{x}_1)^2 \sum(X_2-\bar{x}_2)(Y-\bar{y}) - \sum(X_1-\bar{x}_1)(X_2-\bar{x}_2)\sum(X_1-\bar{x}_1)(Y-\bar{y})}{\sum(X_1-\bar{x}_1)^2 \sum(X_2-\bar{x}_2)^2 - \left[\sum(X_1-\bar{x}_1)(X_2-\bar{x}_2)\right]^2} \qquad (16.13)$$

Apoi, substituind aceste valori în prima ecuaţie a sistemului de ecuaţii normale (după ce a fost împărţită cu n) (16.10), rezultă:

$$a_0 = \bar{y} - a_1\bar{x}_1 - a_2\bar{x}_2 \qquad (16.14)$$

Sistemul (16.9) se poate rezolva utilizînd metodele clasice: metoda simplex, metoda lui Gauss, regula lui Cramer, schema lui Doolittle şi altele.

Vom ilustra în continuare numai unele probleme ale metodei simplex de rezolvare a sistemelor de ecuaţii normale în cazul regresiei multiple liniare.

Sistemul (16.9) se poate scrie sub formă matricolă astfel:

$$\begin{vmatrix} n & \sum X_1 & \sum X_2 & \dots & \sum X_i & & \sum X_k \\ \sum X_1 & \sum X_1^2 & \sum X_2X_1 & \dots & \sum X_iX_1 & \dots & \sum X_kX_1 \\ \sum X_2 & \sum X_1X_2 & \sum X_2^2 & \dots & \sum X_iX_2 & \dots & \sum X_kX_2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \sum X_i & \sum X_1X_i & \sum X_2X_i & \dots & \sum X_i^2 & \dots & \sum X_kX_i \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \sum X_k & \sum X_1X_k & \sum X_2X_k & \dots & \sum X_iX_k & \dots & \sum X_k^2 \end{vmatrix} \cdot \begin{vmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_i \\ \vdots \\ a_k \end{vmatrix} = \begin{vmatrix} \sum Y \\ \sum YX_1 \\ \sum YX_2 \\ \vdots \\ \sum YX_i \\ \vdots \\ \sum YX_k \end{vmatrix} \qquad (16.15)$$

Rezolvarea sistemului (16.15) se face în mai multe etape. In prima etapă se întocmeşte tabelul 16.1, în care în capul liniei şi al coloanelor se înscriu parametrii $a_0, a_1 \dots, a_k$ iar pe ultima coloană şi pe ultimul termenii liberi. La intersecţia ultimului rînd cu ultima coloană se scrie $\sum Y^2$.

Tabelul 16.1. Iteraţia 1.

	a_0	a_1	a_2	...	a_i	...	a_k	
a_0	n	$\sum X_1$	$\sum X_2$	...	$\sum X_i$	...	$\sum X_k$	$\sum Y$
a_1	$\sum X_1$	$\sum X_1^2$	$\sum X_2X_1$	...	$\sum X_iX_1$	...	$\sum X_kX_1$	$\sum YX_1$
a_2	$\sum X_2$	$\sum X_1X_2$	$\sum X_2^2$	...	$\sum X_iX_2$	...	$\sum X_kX_2$	$\sum YX_2$
⋮	....	....	...	...	....	...	.....	....
a_i	$\sum X_i$	$\sum X_1X_i$	$\sum X_2X_i$	...	$\sum X_i^2$	...	$\sum X_kX_i$	$\sum YX_i$
⋮	....	....	...	...	....	...	.....	....
a_k	$\sum X_k$	$\sum X_1X_k$	$\sum X_2X_k$	...	$\sum X_iX_k$	...	$\sum X_k^2$	$\sum YX_k$
	$\sum Y$	$\sum YX_1$	$\sum YX_2$	...	$\sum YX_i$	...	$\sum YX_k$	$\sum Y^2$

In etapa a doua, se alege ca element pivot n de la intersecţia liniei a_0 cu coloana a_0. Apoi, pentru a se obţine iteraţia următoare, se împart elementele de pe primul rînd la elementul pivot şi se obţine astfel prima linie din noul tabel. In continuare, din elementele liniilor 2 ... k se scade prima linie înmulţită cu primul element al liniei respective. (Iteraţia 2).

Iteraţia 2 se mai poate scrie sub forma iteraţiei (2').

Tabelul 16.2 Iteraţie 2

	a_0	a_1	a_2	...	a_i	...	a_k	
a_0	1	$\frac{\sum X_1}{n}$	$\frac{\sum X_2}{n}$	...	$\frac{\sum X_i}{n}$	...	$\frac{\sum X_k}{n}$	$\frac{\sum Y}{n}$
a_1	$\sum X_1 - \frac{n}{n}\sum X_1$	$\sum X_1^2 - \frac{\sum X_1}{n}\sum X_1$	$\sum X_2 X_1 - \frac{\sum X_2}{n}\sum X_1$	...	$\sum X_i X_1 - \frac{\sum X_i}{n}\sum X_1$	...	$\sum X_k X_1 - \frac{\sum X_k}{n}\sum X_1$	$\sum Y X_1 - \frac{\sum Y}{n}\sum X_1$
a_2	$\sum X_2 - \frac{n}{n}\sum X_2$	$\sum X_1 X_2 - \frac{\sum X_1}{n}\sum X_2$	$\sum X_2^2 - \frac{\sum X_2}{n}\sum X_2$	...	$\sum X_i X_2 - \frac{\sum X_i}{n}\sum X_2$	...	$\sum X_k X_2 - \frac{\sum X_k}{n}\sum X_2$	$\sum Y X_2 - \frac{\sum Y}{n}\sum X_2$
..	.....	.....	.....	...	.....	...	.....	.....
a_i	$\sum X_i - \frac{n}{n}\sum X_i$	$\sum X_1 X_i - \frac{\sum X_1}{n}\sum X_i$	$\sum X_2 X_i - \frac{\sum X_2}{n}\sum X_i$	...	$\sum X_i^2 - \frac{\sum X_i}{n}\sum X_i$	...	$\sum X_k X_i - \frac{\sum X_k}{n}\sum X_i$	$\sum Y X_i - \frac{\sum Y}{n}\sum X_i$
..	.....	.....	.....	...	.....	...	.....	.....
a_k	$\sum X_k - \frac{n}{n}\sum X_k$	$\sum X_1 X_k - \frac{\sum X_1}{n}\sum X_k$	$\sum X_2 X_k - \frac{\sum X_2}{n}\sum X_k$	...	$\sum X_i X_k - \frac{\sum X_i}{n}\sum X_i$	...	$\sum X_k^2 - \frac{\sum X_k}{n}\sum X_k$	$\sum Y X_k - \frac{\sum Y}{n}\sum X_k$
	$\sum Y - \frac{n}{n}\sum Y$	$\sum Y X_1 - \frac{\sum X_1}{n}\sum Y$	$\sum Y X_2 - \frac{\sum X_2}{n}\sum Y$	...	$\sum Y X_i - \frac{\sum X_i}{n}\sum Y$	...	$\sum Y X_k - \frac{\sum X_k}{n}\sum Y$	$\sum Y^2 - \frac{\sum Y}{n}\sum Y$

Scanned with OKEN Scanner

Tabelul 16.3. Iteraţia 2'

	a_0	a_1	a_2	...	a_i	...	a_k	
a_0	1	$\bar{x}_1$	$\bar{x}_2$	...	$\bar{x}_i$	...	$\bar{x}_k$	
a_1	0	$\sum(X_1-\bar{X}_1)^2$	$cov(X_2X_1)$	...	$cov(X_iX_1)$	...	$cov(X_kX_1)$	$cov(YX_1)$
a_2	0	$cov(X_1X_2)$	$\sum(X_2-\bar{X}_2)^2$	...	$cov(X_iX_2)$	...	$cov(X_kX_2)$	$cov(YX_2)$
⋮	⋮	.....	.....	...	.....	...	.....	.....
a_i	0	$cov(X_1X_i)$	$cov(X_2X_i)$	...	$\sum(X_i-\bar{X}_i)^2$	...	$cov(X_kX_i)$	$cov(YX_i)$
⋮	⋮	.....	.....	...	.....	...	.....	.....
a_k	0	$cov(X_1X_k)$	$cov(X_2X_k)$	...	$cov(X_iX_k)$	...	$\sum(X_k-\bar{X}_k)^2$	$cov(YX_k)$
	0	$cov(YX_1)$	$cov(YX_2)$	...	$cov(YX_i)$	...	$cov(YX_k)$	$\sum(Y-\bar{Y})^2$

In iteraţia 3 se va alege ca element pivot elementul de la intersecţia coloanei a_1 cu linia a_1, iar calculul elementelor din tabel se va efectua după aceleaşi principii care au stat la baza iteraţiei 2 şi aşa mai departe, pînă cînd matricea parametrilor sistemului de ecuaţii normale se transformă într-o matrice unitate. La ultima iteraţie pe ultima coloană se vor găsi valorile parametrilor ecuaţiei de regresie $a_0, a_1, \ldots, a_i, a_k$.

Pentru ilustrarea metodei de calcul al parametrilor ecuaţiei de regresie mutiplă liniară, considerăm legătura dintre gradul de dezvoltare industrială (Y) a judeţelor din România şi nivelul de înzestrare tehnic a industriei (X_1) şi numărul muncitorilor din industria fiecărui judeţ (X_2). Aceste date se prezintă în tabelul 16.4. Nivelul de dezvoltare industrială a judeţelor ţării se poate caracteriza cu ajutorul indicatorilui - valoarea producţiei globale industriale exprimată în preţurile anului 1963, iar nivelul înzestrării tehnice cu ajutorul indidactorului -ponderea fondurilor fixe industriale ale fiecărui judeţ în totalul fondurilor fixe din industria R.S. România în anul 1975. Deci avem de studiat legătura dintre o variabilă independentă şi două variabile dependente.

Tabelul 16.4. Producţia globală industrială, numărul mediu al muncitorilor şi ponderea judeţului în cadrul ţării după valoarea fondurilor fixe în anul 1975

Judeţul	Producţia globală industrială (miliarde lei preţuri 1963)	Numărul mediu al muncitorilor (mii persoane)	Fonduri fixe industriale (% faţă de total R.S.R.)
Alba	8,1	51,6	1,5
Arad	11,6	60,0	1,2
Argeş	21,9	81,6	5,8
Bacău	16,3	70,1	4,4
Bihor	14,0	76,0	2,1
Bistriţa Năsăud	1,8	13,9	0,4
Botoşani	3,4	17,4	0,5
Braşov	38,3	138,5	5,0
Brăila	11,2	39,9	2,3
Buzău	9,1	35,2	1,4
Caraş Severin	11,5	54,7	2,2
Cluj	20,3	100,8	2,9
Constanţa	12,8	48,3	2,2
Covasna	3,5	21,6	0,6
Dîmboviţa	11,4	61,7	2,9
Dolj	18,1	59,7	3,2
Galaţi	25,4	56,7	5,3
Gorj	5,8	32,9	2,4
Harghita	6,6	46,7	1,0
Hunedoara	21,5	84,1	5,8
Ialomiţa	5,3	16,7	1,1
Iaşi	19,1	57,9	2,5
Ilfov	7,4	39,3	1,4
Maramureş	8,0	53,7	1,4
Mehedinţi	6,4	24,0	1,8
Mureş	17,2	73,1	3,1
Neamţ	15,9	56,7	3,3
Olt	13,5	33,5	2,5
Prahova	32,2	128,1	5,8
Satu Mare	7,7	39,5	0,6
Sălaj	2,1	14,3	0,4
Sibiu	21,3	96,6	2,2
Suceava	9,9	60,4	1,5
Teleorman	7,0	28,0	1,8
Timiş	20,7	103,7	2,3
Tulcea	2,7	20,0	1,3
Vaslui	4,9	22,7	0,8
Vîlcea	5,6	25,8	2,7
Vrancea	4,2	18,6	0,6

Sursa: Rotunjit după Anuarul statistic al Republicii Socialiste România,1976,D.C.S., Bucureşti, p.83-85 şi 58-59.

Se observă, din tabelul 16.4, că municipiul Bucureşti se abate în mare măsură de la nivelul de dezvoltare industrială a judeţelor, atît în privinţa valorii producţiei globale cît şi a numărului de muncitori şi de aceea îl vom exclude din calculele ulterioare. O primă imagine asupra formei ecuaţiei de regresie se va obţine prin reprezentarea grafică a perechilor de valori corelate (Y,X_1), (Y,X_2) şi (X_1,X_2). Aceste legături se prezintă în fig.16.2, fig.16.3 şi fig.16.4.

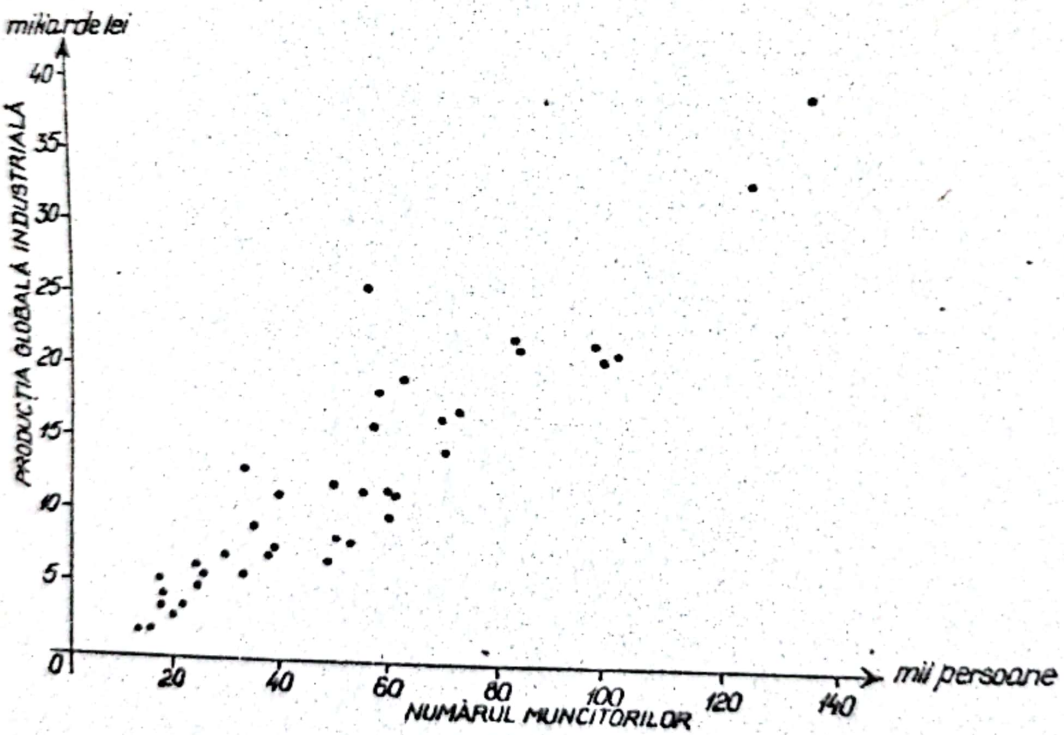


Fig.16.2. LEGĂTURA DINTRE PRODUCȚIA GLOBALĂ INDUSTRIALĂ ȘI NUMĂRUL MUNCITORILOR PE JUDEȚE ÎN ANUL 1975

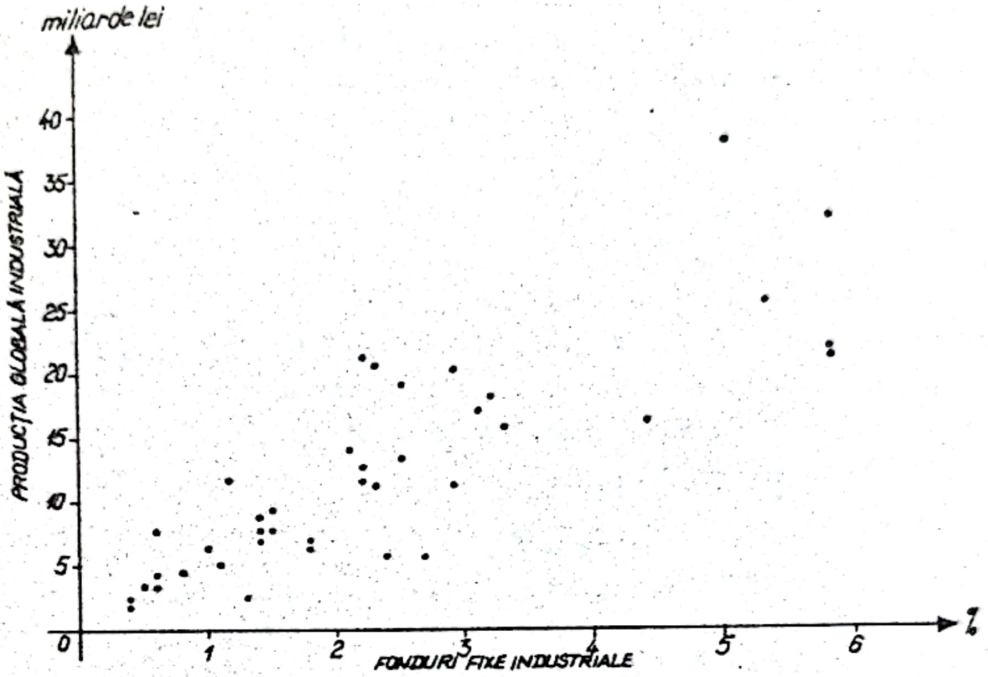


Fig.16.3. LEGĂTURA DINTRE PRODUCȚIA GLOBALĂ INDUSTRIALĂ ȘI VALOAREA FONDURILOR FIXE INDUSTRIALE PE JUDEȚE

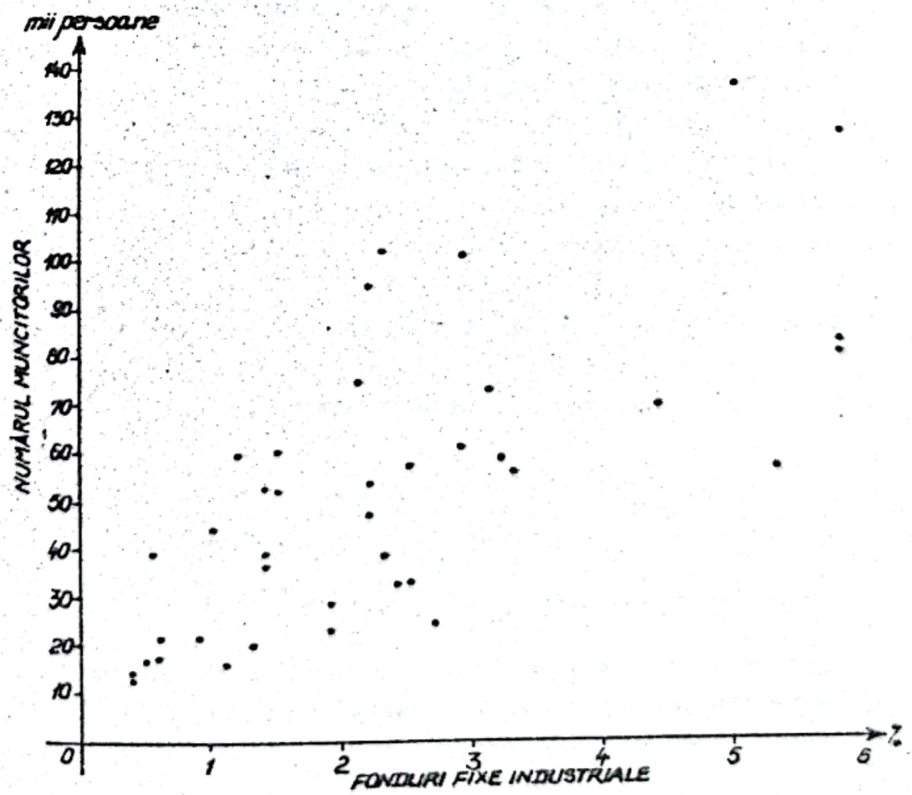


Fig.16.4. LEGĂTURA DINTRE NUMĂRUL MUNCITORILOR DIN INDUSTRIE ȘI VALOAREA FONDURILOR FIXE INDUSTRIALE PE JUDEȚE

Scanned with OKEN Scanner

Din fig. 16.2 se observă că între producţia globală industrială şi numărul muncitorilor din industria judeţelor există o legătură directă, de formă liniară şi de intensitate mare:

$$Y = a + bX_1$$

Interacţiunea dintre cei doi factori cauzali X_1 şi X_2, prezentată în fig. 16.4 se poate descrie tot cu ajutorul unei drepte de regresie, legătura dintre variabile fiind tot directă ca şi în cazurile anterioare, dar de intensitate mai redusă (norul de puncte prezintă o mai mare dispersare în planul X_1OX_2). Dreapta de regresie se poate scrie sub forma ecuaţiei:

$$X_1 = a + bX_2$$

Intrucît între toate perechile de valori există legături simple de tip liniar, înseamnă că şi legătura multiplă de corelaţie este tot liniară şi anume:

$$Y = a_o + a_1X_1 + a_2X_2$$

In această ecuaţie a_o ne arată influenţa tuturor factorilor, în afară de numărul muncitorilor şi gradul de înzestrare cu fonduri fixe industriale, care au acţionat asupra producţiei globăle industriale. Coeficientul de regresie a_1 arată cu cît creşte producţia globală industrială a unui judeţ atunci cînd ponderea fondurilor fixe ale judeţului creşte cu un procent, iar coeficientul a_2 arată că producţia globală a unui judeţ creşte cu a_2 miliarde lei atunci cînd numărul mediu scriptic al muncitorilor din industria judeţului creşte cu 1000 angajaţi.

Sistemul de ecuaţii normale pe baza căruia se pot determina parametrii a_o, a_1 şi a_2, este următorul:

$$\begin{aligned} na_o + a_1 \Sigma X_1 + a_2 \Sigma X_2 &= \Sigma Y \\ a_o \Sigma X_1 + a_1 \Sigma X_1^2 + a_2 \Sigma X_1X_2 &= \Sigma X_1Y \\ a_o \Sigma X_2 + a_1 \Sigma X_1X_2 + a_2 \Sigma X_2^2 &= \Sigma X_2Y \end{aligned}$$

Sumele celor trei variabile, care intervin în acest sistem, se pot determina cu ajutorul tabelului 16.5.

Tabelul 16.5. Calcule preliminare de regresie şi corelaţie multiplă

Judeţul	Prod. globală ind. (miliarde lei) (Y)	Nr. mediu al muncitorilor (mii pers.) (X_1)	Fonduri fixe industr. (%) (X_2)	X_1^2	X_2^2	Y^2	X_1Y	X_2Y	X_1X_2
1	2	3	4	5	6	7	8	9	10
ALBA	8,1	51,0	1,5	2262,46	2,25	65,61	417,96	12,15	77,40
ARAD	11,6	60,0	1,2	3600,00	1,44	134,56	696,00	13,92	72,00
ARGEŞ	21,9	81,6	5,8	6658,56	33,64	479,61	1787,04	127,02	473,28
BACAU	16,3	70,1	4,4	4914,01	19,36	265,69	1142,63	71,72	308,44
BIHOR	14,0	76,0	2,1	5776,00	4,41	196,00	1064,00	29,40	159,60
BISTRITA-NASAUD	1,8	13,9	0,4	193,21	0,16	3,24	25,02	0,72	5,56
BOTOSANI	3,4	17,4	0,5	302,76	0,25	11,56	59,16	1,70	8,70
BRASOV	38,3	138,5	5,0	19182,25	25,00	1466,89	5304,55	191,50	692,50
BRAILA	11,2	39,9	2,3	1592,01	5,29	125,44	446,88	25,76	91,77
BUZAU	9,1	35,2	1,4	1239,04	1,96	82,81	320,32	12,74	49,28
CARAŞ SEVERIN	11,5	54,7	2,2	2992,09	4,84	132,25	629,05	25,30	120,34
CLUJ	20,3	100,8	2,9	10160,64	8,41	412,09	2046,24	58,87	292,32
CONSTANTA	12,8	48,3	2,2	2332,89	4,84	163,84	618,24	28,16	106,26
COVASNA	3,5	21,6	0,6	466,56	0,36	12,25	75,60	2,10	12,96
DIMBOVIŢA	11,4	61,7	2,9	3806,89	8,41	129,96	703,38	33,06	178,93
DOLJ	18,1	59,7	3,2	3564,09	10,24	327,61	1080,57	57,92	191,04
GALATI	25,4	56,7	5,3	3214,89	28,09	645,16	1440,18	134,62	300,51
GORJ	5,8	32,9	2,4	1082,41	5,76	33,64	190,82	13,92	78,96
HARGHITA	3,6	46,7	1,0	2180,89	1,00	43,56	308,22	6,60	46,70
HUNEDOARA	21,5	84,1	5,8	7072,81	33,64	462,25	1808,15	124,70	487,78
IALOMIŢA	5,3	16,7	1,1	278,89	1,21	28,09	88,51	5,83	18,37
IASI	19,1	57,9	2,5	3352,41	6,25	364,81	1105,89	47,75	144,75
ILFOV	7,4	39,3	1,4	1544,49	1,96	54,76	290,82	10,36	55,02
MARAMUREŞ	8,0	53,7	1,4	2883,69	1,96	64,00	429,60	11,20	75,18
MEHEDINŢI	6,4	24,0	1,8	576,00	3,24	40,96	153,60	11,52	43,20
MUREŞ	17,2	73,1	3,1	5343,61	9,61	295,84	1257,32	53,32	226,61
NEAMŢ	15,9	56,7	3,3	3214,89	10,89	252,81	901,53	52,47	187,11
OLT	13,5	33,5	2,5	1122,25	6,25	182,25	452,25	33,75	83,75
PRAHOVA	32,2	128,1	5,8	16409,61	33,64	1036,84	4124,82	186,76	742,98
SATU MARE	7,7	39,5	0,6	1560,25	0,36	59,29	304,15	4,62	23,70
SALAJ	2,1	14,3	0,4	204,49	0,16	4,41	30,03	0,84	5,72
SIBIU	21,3	96,6	2,2	9331,56	4,84	453,69	2057,58	46,86	212,52
SUCEAVA	9,9	60,4	1,5	3648,16	2,25	98,01	597,96	14,85	90,60
TELEORMAN	7,0	28,0	1,8	784,00	3,24	49,00	196,00	12,60	50,40
TIMIŞ	20,7	103,7	2,3	10753,69	5,29	428,49	2146,59	47,61	238,51
TULCEA	2,7	20,0	1,3	400,00	1,69	7,29	54,00	3,51	26,00
VASLUI	4,9	22,7	0,8	515,29	0,64	24,01	111,23	3,92	18,16
VILCEA	5,6	25,8	2,7	665,64	7,29	31,36	144,48	15,12	69,66
VRANCEA	4,2	18,6	0,6	345,96	0,36	17,64	78,12	2,52	11,16
TOTAL	483,7	2064,0	90,2	145929,50	300,48	8687,57	34688,50	1537,29	6077,73

Scanned with OKEN Scanner

Tabelul 16.6. Rezolvarea sistemului de ecuaţii normale prin metoda simplex

Parametrii ecuaţiei de regresie	a_0	a_1	a_2	
a_0	39,00000	2063,99976	90,20000	483,69971
a_1	2063,99976	145929,50000	6077,73047	34688,50000
a_2	90,20000	6077,73047	300,48022	1537,29077
Termenii liberi	483,69971	34688,50000	1537,29077	8687,57031
a_0	1	52,92307	2,31282	12,40256
a_1	0	36696,31250	1304,07031	9089,62500
a_2	0	1304,07031	91,86383	418,58032
Termenii liberi	0	9089,62891	418,58032	2688,45703
a_0	1	0	0,43210	-0,70641
a_1	0	1	0,03554	0,24770
a_2	0	0	45,52132	95,56396
Termeni liberi	0	0	95,56372	436,96875
a_0	1	0	0	-1,61354
a_1	0	1	0	0,17310
a_2	0	0	1	2,09932
Termenii liberi	0	0	0	236,34961

Din ultima iteraţie a tabelului 16.6 rezultă că legătura multiplă dintre variabilele, luate în calcul este descrisă de ecuaţia:

$$Y_{x_1x_2} = -1,161354 + 0,17310\ x_1 + 2,09932\ x_2$$

Aceasta înseamnă că între producţia globală industrială a judeţelor, numărul muncitorilor şi valoarea fondurilor fixe există o legătură directă. Producţia globală industrială creşte în medie cu 0,173 miliarde lei atunci cînd numărul muncitorilor creşte cu 1000 şi cu 2,099 miliarde lei atunci cînd fondurile fixe industria-

le cresc cu 1%. Influenţa tuturor celorlalţi factori în afară de x_1 şi x_2 este negativă.

16.2. Intensitatea corelaţiei multiple liniare

Pentru măsurarea intensităţii corelaţiei multiple liniare se poate utiliza indicatorul numit raport de corelaţie, pe care l-am analizat în cazul legăturilor simple. In condiţiile legăturii dintre k variabile independente $X_1, X_2, \ldots, X_k$ şi o variabilă dependentă Y, raportul de corelaţie se determină cu ajutorul formulei:

$$\eta_{y.x_1 \ldots x_k} = \sqrt{\frac{\sigma^2_{y_{x_1 x_2 \ldots x_k}}}{\sigma^2_y}} \qquad (16.16)$$

sau

$$\eta_{y.x_1 \ldots x_n} = \sqrt{1 - \frac{\sigma^2_{yy_{x_1 x_2 \ldots x_k}}}{\sigma^2_y}} = \sqrt{1 - \frac{\sum(y_j - y_{x_1 x_2 \ldots x_k})^2}{\sum(y_j - \bar{y})^2}} \qquad (16.17)$$

In cazul a două variabile independente, raportul de corelaţie multiplă va fi:

$$\eta_{y.x_1 x_2} = \sqrt{1 - \frac{\sum(y_j - y_{x_i x_2})^2}{\sum(y_j - \bar{y})^2}} \qquad (16.18)$$

Substituind în formula (16.18) datele din tabelul 16.6., rezultă:

$$\eta_{y.x_1 x_2} = \sqrt{1 - \frac{236,34961}{2688,45703}} = \sqrt{0,912088} = 0,955$$

Deci, între producţia globală industrială, pe de o parte, şi numărul mediu al muncitorilor şi valoarea fondurilor fixe pe judeţe, pe de altă parte, există o legătură multiplă liniară intensă.

Scanned with OKEN Scanner

Dacă se cunosc parametrii ecuaţiei de regresie multiplă liniară $a_0, a_1, a_2, \dots, a_k$ atunci calculul raportului de corelaţie se poate simplifica.

Sumele de pătrate ale abaterilor din formula (16.17) se mai pot scrie astfel :

$$\sum(Y_j - Y_{x_1x_2\dots x_k})^2 = \sum(Y_j - Y_{x_1x_2\dots x_k})Y_j$$

$$\sum(Y-Y_{x_1x_2\dots x_k})Y = \sum(Y-a_0-a_1X_1-a_2X_2-\dots-a_kX_k)Y$$

$$\sum(Y-Y_{x_1x_2\dots x_k})^2 = \sum Y^2 - a_0\sum Y - a_1\sum X_1Y - a_2\sum X_2Y - \dots - a_k\sum X_kY \qquad (16.19)$$

iar numitorul

$$\sum(Y_j - \bar{Y})^2 = \sum Y^2 - \frac{(\sum Y)^2}{n} \qquad (16.20)$$

Substituind în relaţia (16.17) pe (16.19) şi (16.20), rezultă:

$$\eta_{y.x_1\dots x_k} = \sqrt{\frac{a_0\sum Y + a_1\sum YX_1 + \dots + a_1\sum YX_1 + \dots a_k\sum YX_k - \frac{(\sum Y)^2}{n}}{\sum Y^2 - \frac{(\sum Y)^2}{n}}} \qquad (16.21)$$

In cazul a două variabile independente X_1 şi X_2 şi o singură variabilă dependentă Y, raportul de corelaţie multiplă liniară devine:

$$\eta_{y.x_1x_2} = \sqrt{\frac{a_0\sum Y + a_1\sum YX_1 + a_2\sum YX_2 - \frac{(\sum Y)^2}{n}}{\sum Y^2 - \frac{(\sum Y)^2}{n}}} \qquad (16.22)$$

Pe baza datelor din tabelele 16.5 şi 16.6 se obţine:

$$\eta_{y.x_1x_2} = \sqrt{\frac{-1,61354.483,7+0,17310.34688,50+2,09932.1537,29 - \frac{483,7^2}{39}}{8687,57 - \frac{483,7^2}{39}}} =$$

$$= \sqrt{0,9121} = 0,955$$

iar în cazul legăturii dintre Y şi X_1, X_2, X_3 raportul de corelaţie

multiplă liniară va fi:

$$\eta_{y.x_1x_2x_3} = \sqrt{\frac{a_0 \sum Y + a_1 \sum YX_1 + a_2 \sum YX_2 + a_3 \sum YX_3 - \frac{(\sum Y)^2}{n}}{\sum Y^2 - \frac{(\sum Y)^2}{n}}} \qquad (16.23)$$

Formula (16.21) şi cazurile sale particulare (16.22) şi (16.23) se utilizează în practică atunci cînd datele empirice, cu privire la variabilele luate în studiu, se prezintă sub forma unor şiruri paralele neponderate.

In cazul al doilea şi al treilea, în cadrul formulelor raportului de corelaţie multiplă, va trebui să se ţină seama de frecvenţele corespunzătoare ale perechilor de valori corelate.

Intensitatea corelaţiei multiple liniare se poate măsura şi cu ajutorul coeficientului de corelaţie multiplă calculat pe baza coeficienţilor de corelaţie simplă dintre variabilele perechi. Astfel, în cazul corelaţiei dintre o variabilă factorială Y şi două variabile independente X_1, X_2 coeficientul de corelaţie multiplă liniară - notat cu $R_{y.x_1x_2}$ -se poate calcula după formula:

$$R_{y.x_1x_2} = \sqrt{\frac{r^2_{yx_1} + r^2_{yx_2} - 2\, r_{yx_1} r_{yx_2} r_{x_1y_2}}{1 - r^2_{x_1x_2}}} \qquad (16.24)$$

în care:

$$r_{yx_1} = \frac{n \sum x_1 y - \sum x_1 \sum y}{\sqrt{\left[n \sum x_1^2 - (\sum x_1)^2\right]\left[n \sum y^2 - (\sum y)^2\right]}}$$

$$r_{yx_2} = \frac{n \sum x_2 y - \sum x_2 \sum y}{\sqrt{\left[n \sum x_2^2 - (\sum x_2)^2\right]\left[n \sum y^2 - (\sum y)^2\right]}}$$

$$r_{x_1x_2} = \frac{n \sum x_1 x_2 - \sum x_1 \sum x_2}{\sqrt{\left[n \sum x_1^2 - (\sum x_1)^2\right]\left[n \sum x_2^2 - (\sum x_2)^2\right]}}$$

Calculînd coeficientul de corelaţie multiplă liniară, pe baza formulei (16.24), rezultă:

$$r_{yx_1} = \frac{39.34688,5 - 2064,0.483,7}{\sqrt{(39.145929,5 - 2064,0^2)(39.8687,57 - 483,7^2)}} = 0,9151$$

$$r_{yx_2} = \frac{39.1537,29 - 90,2.483,7}{\sqrt{(39.300,48 - 90,2^2)(39.8687,57 - 483,7^2)}} = 0,8423$$

$$r_{x_1x_2} = \frac{39.6077,73 - 2064,0.90,2}{\sqrt{(39.145929,5 - 2064,0^2)(39.300,48 - 90,2^2)}} = 0,7103$$

iar

$$R_{y.x_1x_2} = \sqrt{\frac{0,9151^2 + 0,8423^2 - 2.0,9151.0,8423.0,7103}{1 - 0,7103^2}} = 0,955$$

Coeficientul de corelaţie multiplă liniară se poate determina şi cu ajutorul unei alte formule bazată pe determinanţii compuşi din coeficienţii de corelaţie simplă liniară, astfel:

- în cazul a trei variabile:

$$R_{y.x_1x_2} = \sqrt{1 - \frac{\Delta yy_x}{\Delta_y}} \qquad (16.25)$$

de unde

$$\Delta yy_x = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{x_1y} & 1 & r_{x_1x_2} \\ r_{x_2y} & r_{x_1x_2} & 1 \end{vmatrix} \quad \text{şi} \quad \Delta_y = \begin{vmatrix} 1 & r_{x_1x_2} \\ r_{x_1x_2} & 1 \end{vmatrix}$$

Pentru aplicarea formulei (16.25), la exemplul din tabelul 16.4, se calculează:

$$\Delta_{yy_x} = \begin{vmatrix} 1 & 0,91513 & 0,84228 \\ 0,91513 & 1 & 0,71026 \\ 0,84228 & 0,71026 & 1 \end{vmatrix} = 0,04356$$

si

$$\Delta_y = \begin{vmatrix} 1 & 0,71026 \\ 0,71026 & 1 \end{vmatrix} = 0,49553$$

iar

$$R_{y.x_1x_2} = \sqrt{1 - \frac{0,04356}{0,49553}} = 0,955$$

-în cazul a patru variabile:

$$R_{y.x_1x_2x_3} = \sqrt{1 - \frac{\Delta_{yy_x}}{\Delta_y}} \qquad (16.26)$$

unde

$$\Delta_{yy_x} = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & r_{yx_3} \\ r_{x_1y} & 1 & r_{x_1x_2} & r_{x_1x_3} \\ r_{x_2y} & r_{x_2x_1} & 1 & r_{x_2x_3} \\ r_{x_3y} & r_{x_3x_1} & r_{x_3x_2} & 1 \end{vmatrix}$$

şi

$$\Delta_y = \begin{vmatrix} 1 & r_{x_1x_2} & r_{x_1x_3} \\ r_{x_2x_1} & 1 & r_{x_2x_3} \\ r_{x_3x_1} & r_{x_3x_2} & 1 \end{vmatrix}$$

- în cazul a k+1 variabile:

$$R_{y_{x_1x_2\ldots x_k}} = \sqrt{1 - \frac{\Delta_{yy_{x_1x_2\ldots x_k}}}{\Delta_y}} \qquad (16.27)$$

unde

$$\Delta_{yyx_1x_2\ldots x_k} = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} \ldots & r_{yx_i} \ldots & r_{yx_k} \\ r_{x_1y} & 1 & r_{x_1x_2} \ldots & r_{x_1x_i} \ldots & r_{x_1x_k} \\ r_{x_2y} & r_{x_2x_1} & 1 \ldots & r_{x_2x_i} \ldots & r_{x_2x_k} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ r_{x_ky} & r_{x_kx_1} & r_{x_kx_2} \ldots & r_{x_kx_i} \ldots & 1 \end{vmatrix}$$

şi

$$\Delta_y = \begin{vmatrix} 1 & r_{x_1x_2} & r_{x_1x_3} \ldots r_{x_1x_i} \ldots & r_{x_1x_k} \\ r_{x_2x_1} & 1 & r_{x_2x_3} \ldots r_{x_2x_i} \ldots & r_{x_2x_k} \\ r_{x_3x_1} & r_{x_3x_2} & 1 \ldots r_{x_3x_i} \ldots & r_{x_3x_k} \\ \ldots & \ldots & \ldots & \ldots \\ r_{x_kx_1} & r_{x_kx_2} & r_{x_kx_3} \ldots r_{x_kx_i} \ldots & 1 \end{vmatrix}$$

16.2.3. Corelaţia parţială

In cazul corelaţiei simple se studiază legătura liniară dintre doi factori, neglijîndu-se influenţa celorlalţi factori care acţionează asupra variabilei rezultative Y, în timp ce în cazul corelaţiei multiple liniare se studiază influenţa simultană a două sau mai multe caracteristici factoriale asupra caracteristicii rezulative.

O altă problemă ce se poate rezolva cu ajutorul metodei regresiei şi corelaţiei este determinarea influenţei unei variabile independente asupra variabilei dependente, presupunînd că celelalte variabile independente se menţin la un nivel constant. Indicatorul intensităţii unei astfel de corelaţii poartă denumi-

rea de coeficient de corelatie partială.

Coeficienţii de corelaţie parţială, în cazul a două variabile independente şi o singură variabilă dependentă, se calculează cu ajutorul formulelor:

- între Y şi X_1 neglijînd influenţa lui X_2 :

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \, r_{x_1x_2}}{\sqrt{(1-r^2_{yx_2})} \sqrt{(1-r^2_{x_1x_2})}} \qquad (16.28)$$

- între Y şi X_2 neglijînd influenţa lui X_1 :

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \, r_{x_1x_2}}{\sqrt{(1-r^2_{yx_1})} \sqrt{(1-r^2_{x_1x_2})}} \qquad (16.29)$$

Coeficienţii de corelaţie parţială (16.28) şi (16.29) poartă denumirea de coeficienţi de corelaţie parţială de ordinul întîi.

In exemplul din tabelul 16.4 coeficienţii de corelaţie parţială vor avea următoarele valori:

$$r_{yx_1 \cdot x_2} = \frac{0,9151 - 0,8423.0,7103}{\sqrt{(1-0,8423^2)} \sqrt{(1-0,7103^2)}} = 0,8351$$

$$r_{yx_2 \cdot x_1} = \frac{0,2423 - 0,9151.0,7103}{\sqrt{(1-0,9151^2)} \sqrt{(1-0,7103^2)}} = 0,67758$$

In cazul a patru variabile Y, X_1, X_2, X_3 coeficienţii de corelaţie parţială de ordinul al doilea, se pot determina cu ajutorul coeficienţilor de corelaţie parţială de ordinul întîi, după formulele:

- între Y şi X_1 neglijînd influenţa lui X_2 şi X_3 :

$$r_{yx_1 \cdot x_2x_3} = \frac{r_{yx_1 \cdot x_2} - r_{yx_3 \cdot x_2} \, r_{x_1x_3 \cdot x_2}}{\sqrt{(1-r^2_{yx_3 \cdot x_2})} \sqrt{(1-r^2_{x_1x_3 \cdot x_2})}} \qquad (16.30)$$

- între Y şi X_2 neglijînd influenţa lui X_1 şi X_3 :

$$r_{yx_2 \cdot x_1 x_3} = \frac{r_{yx_2 \cdot x_1} - r_{yx_3 \cdot x_1} \cdot r_{x_2 x_3 \cdot x_1}}{\sqrt{(1-r^2_{yx_3 \cdot x_1})}\ \sqrt{(1-r^2_{x_2 x_3 \cdot x_1})}} \qquad (16.31)$$

- între Y şi X_3 neglijînd influenţa lui X_1 şi X_2 :

$$r_{yx_3 \cdot x_1 x_2} = \frac{r_{yx_3 \cdot x_1} - r_{yx_2 \cdot x_1} \cdot r_{x_2 x_3 \cdot x_1}}{\sqrt{(1-r^2_{yx_2 \cdot x_1})}\ \sqrt{(1-r^2_{x_2 x_3 \cdot x_1})}} \qquad (16.32)$$

In cazul general:

$$r_{yx_1 \cdot x_2 x_3 \ldots x_k} = \frac{r_{yx_1 \cdot x_2 x_3 \ldots x_{k-1}} - r_{yx_k \cdot x_2 x_3 \ldots x_{k-1}} \cdot r_{x_1 x_k \cdot x_2 x_3 \ldots x_{k-1}}}{\sqrt{(1-r^2_{yx_k \cdot x_2 x_3 \ldots x_{k-1}})}\ \sqrt{(1-r^3_{x_1 x_k \cdot x_2 x_3 \ldots x_{k-1}})}}$$

Calculul indicatorilor de regresie şi corelaţie multiplă şi parţială, chiar în cazul legăturilor liniare este foarte complicat şi necesită utilizarea calculatoarelor electronice. Volumul calculelor şi gradul de complexitate cresc odată cu creşterea numărului variabilelor corelate.

16.2.4. Determinaţie multiplă şi parţială

Utilizarea metodei corelaţiei în statistică oferă posibilitatea determinării contribuţiei fiecărei variabile independente asupra variaţiei varibilei dependente ca şi influenţa simultană a tuturor variabilelor independente asupra variabilei dependente. Gradul de influenţă, în primul caz, se determină cu ajutorul coeficientilor de determinatie parţială, iar în cel de al doilea caz cu coeficientul de determinatie multiplă.

Coeficientul de determinaţia multiplă se poate calcula prin ridicarea la pătrat a coeficientului sau raportului corelaţiei multiple.

In cazul a trei variabile Y, X_1, X_2, coeficientul de determinaţie se poate obţine pe baza coeficientului de corelaţie multiplă, calculat cu ajutorul coeficienţilor de corelaţie simplă:

$$R^2_{y.x_1x_2} = \frac{r^2_{yx_1} + r^2_{yx_2} - 2\, r_{yx_1}\, r_{yx_2}\, r_{x_1x_2}}{1 - r^2_{x_1x_2}} \qquad (16.34)$$

Din formula (16.34) rezultă că coeficientului determinaţiei multiple este mai mare decît oricare dintre coeficienţii determinaţiei simple:

$$R^2_{y.x_1x_2} > r^2_{yx_1}$$

şi

$$R^2_{y.x_1x_2} > r^2_{yx_2}$$

Coeficientul determinaţiei multiple $R^2_{y.x_1x_2}$ arată care este influenţa simultană a factorilor cauzali (luaţi în calcul) asupra fenomenului efect şi variază între limitele 0 şi 1.

$$0 < R^2_{y.x_1x_2} < 1$$

Pentru a se interpreta mai uşor coeficientul determinaţiei multiple se poate exprima în procente.In ac est caz el va arăta cît la sută din variaţia variabilei rezultative se datorează variaţiei variabilelor factoriale.

Aplicînd formula (16.34) la exemplul din tabelul 16.4, rezultă:

$$R^2_{y.x_1x_2} = \frac{0{,}9151^2 + 0{,}8423^2 - 2.0{,}9151.0{,}8423.0{,}7103}{1 - 0{,}7103^2} = 0{,}9121$$

Aceasta înseamnă că variaţia producţiei globale industriale este determinată în proporţie de 91,21% de variaţia numărului mediu de muncitori şi de valoarea fondurilor fixe pe judeţe şi numai 8,79% se datorează influenţei altor factori decît X_1 şi X_2.

Dacă Y este independent de X_1 şi X_2, atunci coeficientul determinaţiei multiple este nul: $R^2_{y.x_1x_2} = 0$, cînd $r_{yx_1} = 0$ şi $r_{yx_2} = 0$.

In cazul legăturilor funcţionale între Y, X_1 şi X_2 coeficientul determinaţiei totale este egal cu 1. Aceasta înseamnă că $r_{x_1x_2} = 0$, iar formula (16.34) devine:

$$R^2_{y.x_1x_2} = r^2_{yx_1} + r^2_{yx_2} \qquad (16.35)$$

Coeficienţii de determinaţie parţială se pot calcula pe baza formulelor următoare :

- între Y şi X_1 neglijînd pe X_2 :

$$r^2_{yx_1.x_2} = \frac{R^2_{y.x_1x_2} - r^2_{yx_2}}{1 - r^2_{yx_2}} \qquad (16.36)$$

- între Y şi X_2 neglijînd pe X_1 :

$$r^2_{yx_2.x_1} = \frac{R^2_{y.x_1x_2} - r^2_{yx_1}}{1 - r^2_{yx_1}} \qquad (16.37)$$

Prin aplicarea formulelor (16.36) şi (16.37) la exemplul din tabelul 16.4, se obţine:

$$r^2_{yx_1.x_2} = \frac{0,9121 - 0,8423^2}{1 - 0,8423^2} = 0,69745$$

şi

$$r^2_{yx_2.x_1} = \frac{0,9121 - 0,9151^2}{1 - 0,9151^2} = 0,45938$$

Calculul coeficienţilor de determinaţie parţială se poate face şi pe baza coeficienţilor de corelaţie parţială (prin ridicarea la pătrat a acestora), cu ajutorul formulei:

$$r^2_{yx_1.x_2} = (r_{yx_1.x_2})^2 = \frac{r_{yx_1} - r_{yx_2}\, r_{x_1x_2}}{(1-r^2_{yx_2})(1-r^2_{x_1x_2})} \qquad (16.38)$$

şi

$$r^2_{yx_2.x_1} = (r_{yx_2.x_1})^2 = \frac{r_{yx_2} - r_{yx_1}\, r_{x_1x_2}}{(1-r^2_{yx_1})(1-r^2_{x_1x_2})} \qquad (16.39)$$

Desigur, formulele (16.36 şi (16.37) sînt echivalente cu (16.30) şi respectiv (16.39).

Scanned with OKEN Scanner

Dacă, pe lîngă coeficienţii de corelaţie parţială, se cunosc şi mediile aritmetice şi abaterile standard ale variabilelor Y, X_1 şi X_2, atunci se pot determina pe această bază şi coeficienţii ecuaţiei regresiei multiple, cu ajutorul formularelor:

$$a_1 = r_{yx_1 \cdot x_2} \frac{\sigma_{y \cdot x_1 x_2}}{\sigma_{x_1 \cdot yx_2}} \qquad (16.40)$$

$$a_2 = r_{yx_2 \cdot x_1} \frac{\sigma_{y \cdot x_1 x_2}}{\sigma_{x_2 \cdot yx_1}} \qquad (16.41)$$

unde

$$\sigma_{y \cdot x_1 x_2} = \sqrt{\sigma_y^2 (1 - r_{yx_2}^2)(1 - r_{yx_1}^2)}$$

$$\sigma_{y_2 \cdot yx_1} = \sqrt{\sigma_{x_1}^2 (1 - r_{x_1 x_2}^2)(1 - r_{yx_1 \cdot x_2}^2)}$$

$$\sigma_{x_2 \cdot yx_1} = \sqrt{\sigma_{x_2}^2 (1 - r_{x_1 x_2}^2)(1 - r_{yx_2 \cdot x_1}^2)}$$

Coeficientul determinaţiei multiple se poate calcula şi cu ajutorul raportuluide corelaţie multiplă liniară .Astfel ,pentru trei variabile, se poate porni de la formula (16.17) şi se obţine:

$$R_{y \cdot x_1 x_2}^2 = \eta_{y \cdot x_1 x_2}^2 = 1 - \frac{\sum (Y_j - Y_{x_1 x_2})^2}{\sum (Y_j - \bar{y})^2} = D_{y, x_1 x_2} \qquad (16.42)$$

sau

$$R_{y \cdot x_1 x_2}^2 = 1 - \frac{a_0 \sum Y + a_1 \sum YX_1 + a_2 \sum YX_1 + a_2 \sum YX_2 - \frac{(\sum Y)^2}{n}}{\sum Y^2 - \frac{(\sum Y)^2}{n}} \qquad (16.43)$$

Parametrul a_0 rezultă imediat din prima ecuaţie a sistemului de ecuaţii normale în care s-au substituit coeficienţii a_1 şi a_2 deja cunoscuţi:

$$a_0 = \frac{\sum Y - a_1 \sum X_1 - a_2 \sum X_2}{n} = \bar{Y} - a_1 \bar{x}_1 - a_2 \bar{x}_2 \qquad (16.44)$$

Scanned with OKEN Scanner

Coeficienţii determinaţiei parţiale se pot obţine din formula (16.43). Substituind pe a_0 din relaţia (16.44) în (16.43) rezultă:

$$R^2_{y.x_1x_2} = \frac{\bar{Y}\sum X - a_1\bar{x}_1\sum Y - a_2\bar{x}_2\sum Y + a_1\sum YX_1 + a_2\sum YX_2 - \frac{(\sum Y)^2}{n}}{\sum Y^2 - \frac{(\sum Y)^2}{n}} =$$

$$= \frac{a_1(\sum YX_1 - \frac{\sum X_1\sum Y}{n}) + a_2(\sum YX_2 - \frac{\sum X_2\sum Y)}{n}}{\sum Y^2 - \frac{(\sum Y)^2}{n}}$$

Separînd numitorii, se obţine:

$$R^2_{y.x_1x_2} = \frac{a_1(n\sum YX_1 - \sum X_1\sum Y)}{n\sum Y^2 - (\sum Y)^2} + \frac{a_2(n\sum YX_2 - \sum X_2\sum Y)}{n\sum Y^2 - (\sum Y)^2} \qquad (16.45)$$

In formula (16.45) termenii din membrul al doilea reprezintă cei doi coeficienţi de determinaţie parţială:

- între Y şi X_1:

$$r^2_{y\,x_1x_2} = \frac{a_1(n\sum YX_1 - \sum X_1\sum Y)}{n\sum Y^2 - (\sum Y)^2} = d_{yx_1.x_2} \qquad (16.46)$$

- între Y şi X_2:

$$r^2_{yx_2.x_1} = \frac{a_2(n\sum YX_2 - \sum X_1\sum Y)}{n\sum Y^2 - (\sum Y)^2} = d_{yx_2.x_1} \qquad (16.47)$$

Deci

$$R^2_{y.x_1x_2} = r^2_{yx_1.x_2} + r^2_{yx_2.x_1} \qquad (16.48)$$

Din relaţia (16.48) rezultă că suma coeficienţilor determinaţiei parţiale este egală cu coeficientul determinaţiei multiple liniare.

Pe măsură ce în modelul de corelaţie se adaugă noi variabile, coeficientul corelaţiei multiple liniare (ca şi coeficientul determinaţiei multiple) creşte şi se pierde cîte un grad de libertate a variaţiei explicate. Altfel spus, se obţine o tot mai bună estimare a fenomenului efect Y. Creşterea numărului de variabile în modelul de regresie determină o creştere mai mult decît proporţională a volumului calculelor. Astfel, în cazul unei probleme de regresie şi corelaţie multiplă dintre patru variabile Y, X_1, X_2, X_3, volumul calculelor se măreşte cu mai mult de o treime faţă de aceeaşi problemă în cazul utilizării a trei variabile Y, X_1 şi X_2.

In cazul exemplului din tabelul 16.4, calculul coeficienţilor de determinaţie pe baza formulelor (16.46) şi (16.47) se prezintă astfel:

$$d_{yx_1 \cdot x_2} = \frac{0,71310(39.34688 - 2064.483,7)}{39.8687,57 - 483,7^2} = 0,58$$

şi

$$d_{yx_2 \cdot x_1} = \frac{2,09932(39.1557,29 - 90,2.483,7)}{39.8687,57 - 483,7^2} = 0,34$$

iar $D_{y.x_1x_2} = 0,58 + 0,34 = 0,92$

In cazul a patru variabile Y, X_1, X_2, X_3 coeficientul determinaţiei totale va fi egal cu:

$$D_{y.x_1x_2x_3} = \eta^2_{y.x_1x_2x_3} = 1 - \frac{\sum(Y_i - \bar{Y}_{x_1x_2})^2}{\sum(Y_i - \bar{y})^2}$$

$$D_{y.x_1x_2x_3} = 1 - \frac{a_0\sum Y + a_1\sum YX_1 + a_2\sum YX_2 + a_3\sum YX_3 - \frac{(\sum Y)^2}{n}}{\sum Y^2 - \frac{(\sum Y)^2}{n}} \qquad (16.49)$$

Pentru stabilirea formulelor de calcul ale celor trei coeficienţi de determinaţie parţială, vom substitui în (16.49) pe a_0 cu valoarea lui din prima ecuaţie a sistemului de ecuaţii normale:

$$a_0 = \bar{y} - a_1\bar{x}_1 - a_2\bar{x}_2 - a_3\bar{x}_3$$

deci

$$D_{y.x_1x_2x_3} = \frac{\bar{y}\sum Y - a_1\bar{x}_1\sum Y - a_2\bar{x}_2\sum Y - a_3\bar{x}_3\sum Y + a_1\sum YX_1 + a_2\sum YX_2 + a_3\sum YX_3 - \frac{(\sum Y)^2}{n}}{\sum Y^2 - \frac{(\sum Y)^2}{n}}$$

$$= \frac{a_1(\sum YX_1 - \frac{\sum X_1\sum Y}{n}) + a_2(\sum YX_2 - \frac{\sum X_2\sum Y}{n}) + a_3(\sum YX_3 - \frac{\sum X_3\sum Y}{n})}{\sum Y^2 - \frac{(\sum Y)^2}{n}}$$

Rezultă:

$$d_{yx_1 \cdot x_2x_3} = \frac{a_1(n\sum YX_1 - \sum X_1 \sum Y)}{n\sum Y^2 - (\sum Y)^2} \tag{16.50}$$

$$d_{yx_2 \cdot x_1x_3} = \frac{a_2(n\sum YX_2 - \sum X_2 \sum Y)}{n\sum Y^2 - (\sum Y)^2} \tag{16.51}$$

$$d_{yx_3 \cdot x_1x_2} = \frac{a_3(n\sum YX_3 - \sum X_3 \sum Y)}{n\sum Y^2 - (\sum Y)^2} \tag{16.52}$$

In general, pentru k variabile coeficienţii de determinaţie se vor calcula astfel:

$$D_{y \cdot x_1x_2 \ldots x_k} = 1 - \frac{a_0\sum Y + a_1\sum YX_1 + a_2\sum YX_2 + \ldots + a_i\sum YX_i + \ldots + a_k\sum YX_k - \frac{(\sum Y)^2}{n}}{\sum Y^2 - \frac{(\sum Y)^2}{n}} \tag{16.53}$$

şi

$$d_{yx_1 \cdot x_2x_3 \ldots x_k} = \frac{a_1(n\sum YX_1 - \sum X_1 \sum Y)}{n\sum Y^2 - (\sum Y)^2} \tag{16.54}$$

$$d_{yx_2 \cdot x_1x_3x_4 \ldots x_k} = \frac{a_2(n\sum YX_2 - \sum X_2 \sum Y)}{n\sum Y^2 - (\sum Y)^2} \tag{16.55}$$

$$d_{yx_3 \cdot x_1x_2x_4 \ldots x_k} = \frac{a_3(n\sum YX_3 - \sum X_3 \sum Y)}{n\sum Y^2 - (\sum Y)^2} \tag{16.56}$$

.....

$$d_{yx_1 \cdot x_1x_2 \ldots x_k} = \frac{a_1(n\sum YX_1 - \sum X_1 \sum Y)}{n\sum Y^2 - (\sum Y)^2} \tag{16.57}$$

.....

$$d_{yx_k \cdot x_1x_2 \ldots x_{k-1}} = \frac{a_k(n\sum YX_k - \sum X_k \sum Y)}{n\sum Y^2 - (\sum Y)^2} \tag{16.58}$$

Intre coeficientul determinaţiei totale şi coeficienţii determinaţiei parţiale există următoarea relaţie: suma coeficienţilor de determinaţie parţială este egală cu coeficientul determinaţiei totale.

$$D_{y.x_1x_2\ldots x_k} = d_{yx_1.x_2x_3\ldots x_k} + d_{yx_2.x_1x_3\ldots x_k} + \ldots + d_{yx_1.x_1x_2\ldots x_k} + \ldots + d_{yx_k.x_1x_2\ldots x_{k-1}}$$

sau

$$D_{y.x_1x_2x_3\ldots x_k} = \sum_{i=1}^{k} d_{yx_i.x_1x_2\ldots x_k} \qquad (16.59)$$

Relaţia (16.59) evidenţiază faptul că variaţia totală a fenomenului efect Y se poate descompune pe factori cauzali de influenţă $X_1X_2\ldots X_k$.

16.3. Regresie şi corelaţie multiplă curbilinie

16.3.1. Regresie multiplă curbilinie

Am văzut anterior că regresia multiplă liniară se caracterizează prin faptul că fenomenul efect variază în mod uniform atunci cînd fenomenele cauzale se modifică cu o unitate, indiferent de mărimea fiecărei variabile independente. Pentru majoritatea fenomenelor dezvoltarea lor nu se face după modelul liniar, ci după o curbă de un anumit tip. In dezvoltarea unor fenomene se poate identifica un nivel optim de creştere, care corespunde unor anumite niveluri ale variabilelor cauzale. Modeul liniar nu poate descrie o astfel de evoluţie şi de aceea este necesar să se recurgă la modele curbilinii.

In forma cea mai simplă, o problemă de regresie şi corelaţie multiplă curbilinie se formulează astfel: fie o variabilă dependentă Y care este influenţată de K variabile cauzale, independente $X_1, X_2, \ldots, X_k$, iar legătura dintre variabile este de forma:

$$Y = a + f_1(X_1) + f_2(X_2) + \ldots + f(X_i) + \ldots + f_k(X_k) \qquad (16.60)$$

Relaţia (16.60) arată că variaţia fenomenului efect Y este rezultatul unei sume de funcţii de X_1, de $X_2, \ldots$, de X_k şi de alţi factori în afară de X_i $(i = 1,2,\ldots,k)$ a căror influenţă totală este egală cu a_0 . Desigur, o parte din funcţiile $f(X_i)$, $(i = 1,2,\ldots,k)$ pot să fie liniare, iar o altă parte curbilinii.

Pentru ca legătura multiplă de corelaţie să fie curbilinie este suficient ca una din funcţiile $f(X_i)$ să fie de formă curbilinie.

Determinarea parametrilor ecuaţiei regresiei multiple curbilinii se poate face cu ajutorul metodei celor mai mici pătrate sau pe baza metodei aproximării grafice succesive.

Alegerea modelului regresiei multiple curbilinii este necesar să se exprime printr-o funcţie, cît mai simplă din punct de vedere matematic, în aşa fel încît parametrii ecuaţiei de regresie să se poată determina cu ajutorul unor metode pe care le-am prezentat în partea anterioară a acestui capitol. Dacă se utilizează un model de forma polinomului de gradul doi, atunci ecuaţia regresiei multiple curbilinii (16.60) va deveni:

$$Y=a+a_1X_1+b_1X_1^2+a_2X_2+b_2X_2^2+\ldots+a_iX_i+b_iX_i^2+\ldots+a_kX_k+b_xX_k^2 \qquad (16.61)$$

în care:

$$f(X_i) = a_0+a_iX_i + b_iX_i^2 \qquad i = 1,2, \ldots , k$$

In practica statistică se pot distinge două situaţii diferite:

1) cînd se cunosc, sau se pot stabili, pe baza analizei statistice, ori există anumite studii, efectuate anterior asupra acelorași variabile, formele ecuaţiilor simple de regresie dintre variabile efect şi fiecare variabilă cauzală, se poate determina o funcţie de regresie multiplă curbilinie corespunzătoare ai căror parametri se estimează cu ajutorul metodei celor mai mici pătrate;

2) dacă nu se cunoaşte forma reală a legăturii dintre variabila dependentă şi fiecare variabilă independentă, atunci se procedează la o reprezentare grafică cît mai adecvată a legăturii simple şi, analizînd esenţa fenomenelor corelate prin încercări succesive, se ajunge la stabilirea celei mai corespunzătoare forme a ecuaţiei de regresie simplă curbilinie şi apoi se va elabora modelul regresiei multiple curbilinii.

Parametrii ecuaţiei regresiei multiple (16.61) de tip polinomial se pot determina, pe baza aplicării metodei celor mai mici pătrate, cu ajutorul sistemului de ecuaţii normale:

$$n\,a+a_1\sum X_1+b_1\sum X_1^2+a_2\sum X_2+b_2\sum X_2^2+\ldots+a_i\sum X_i+b_i\sum X_i^2+\ldots+$$
$$+a_k\sum X_k+b_k\sum X_k^2=\sum Y_j$$

$$a\sum X_1+a_1\sum X_1^2+b_1\sum X_1^3+a_2\sum X_1X_2+b_2\sum X_1X_2^2+\ldots+a_i\sum X_1X_i+b_i\sum X_1X_i^2+\ldots+$$
$$+a_k\sum X_1X_k+b_k\sum X_1X_k^2=\sum X_1Y_j$$

$$a\sum X_1^2+a_1\sum X_1^3+b_1\sum X_1^4+a_2\sum X_1^2X_2+b_2\sum X_2^2+\ldots+a_i\sum X_1^2X_i+b_i\sum X_1^2X_i^2+\ldots+$$
$$+a_k\sum X_1^2X_k+b_k\sum X_1^2X_k^2=\sum X_1^2Y_j \qquad (16.62)$$

$$a\sum X_2+a_1\sum X_1X_2+b_1\sum X_1^2X_2+a_2\sum X_2^2+b_2\sum X_2^3+\ldots+a_i\sum X_2X_i+b_i\sum X_1^2X_2+$$
$$+\ldots+a_k\sum X_2X_k+b_k\sum X_2X_k^2=\sum X_2Y_j$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$a\sum X_i+a_1\sum X_1X_i+b_1\sum X_1^2X_i+a_2\sum X_2X_i+b_2\sum X_2^2X_i+\ldots+a_i\sum X_i^2+b_i\sum X_i^3+$$
$$+\ldots+a_k\sum X_iX_k+b_k\sum X_iX_k^2=\sum X_iY_j$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$a\sum X_k+a_1\sum X_1X_k+b_1\sum X_1^2X_k+a_2\sum X_2X_k+b_2\sum X_2^2X_k+\ldots+a_i\sum X_iX_k+b_i\sum X_i^2X_k+$$
$$+\ldots+a_k\sum X_k^2+b_k\sum X_k^3=\sum X_kY_j$$

Pentru a prezenta mai detaliat metodele de regresie şi corelaţie multiplă curbilinie, să considerăm cazul mai simplu al legăturii de tip parabolic între două variabile independente X_1 şi X_2 şi o singură variabilă dependentă Y. Conform relaţiei (16.61) modelul regresiei multiple curbilinii va fi descris de următoarea ecuaţie:

$$Y = a + a_1X_1+b_1X_1^2 + a_2X_2 + b_2X_2^2 \qquad (16.63)$$

Parametrii acestei ecuaţii de regresie se pot determina tot cu ajutorul metodei celor mai mici pătrate, astfel:

$$S = \sum(Y_j - Y_{x_1x_2})^2 = \text{minim}$$

sau

$$S = \sum(Y_j - a - a_1X_1 - b_1X_1^2 - a_2X_2 - b_2X_2^2)^2 = \text{minim}$$

$$\frac{\partial S}{\partial a} = n\,a + a_1\sum X_1 + b_1\sum X_1^2 + a_2\sum X_2 + b_2\sum X_2^2 - \sum Y_j = 0 \qquad (16.64)$$

$$\frac{\partial S}{\partial a_1} = a\sum X_1 + a_1\sum X_1^2 + b_1\sum X_1^3 + a_2\sum X_1X_2 + b_2\sum X_1X_2^2 - \sum YX_1 = 0$$

$$\frac{\partial S}{\partial b_1} = a\sum X_1^2 + a_1\sum X_1^3 + b_1\sum X_1^4 + a_2\sum X_1^2X_2 + b_2\sum X_1^2X_2^2 - \sum YX_1^2 = 0$$

$$\frac{\partial S}{\partial a_2} = a\sum X_2 + a_1\sum X_1X_2 + b_1\sum X_1^2X_2 + a_2\sum X_2^2 + b_2\sum X_2^3 - \sum YX_2 = 0$$

$$\frac{\partial S}{\partial b_2} = a\sum X_2^2 + a_1\sum X_1X_2^2 + b_1\sum X_1^2X_2^2 + a_2\sum X_2^3 + b_2\sum X_2^4 - \sum YX_2^2 = 0$$

Prin transformări elementare, sistemul (16.64) devine:

$$\begin{aligned}
&n\,a + a_1\sum X_1 + b_1\sum X_1^2 + a_2\sum X_2 + b_2\sum X_2^2 = \sum Y\\
&a\sum X_1 + a_1\sum X_1^2 + b_1\sum X_1^3 + a_2\sum X_1X_2 + b_2\sum X_1X_2^2 = \sum YX_1\\
&a\sum X_1^2 + a_1\sum X_1^3 + b_1\sum X_1^4 + a_2\sum X_1^2X_2 + b_2\sum X_1^2X_2^2 = \sum YX_1^2\\
&a\sum X_2 + a_1\sum X_1X_2 + b_1\sum X_1^2X_2 + a_2\sum X_2^2 + b_2\sum X_2^3 = \sum YX_2\\
&a\sum X_2^2 + a_1\sum X_1X_2^2 + b_1\sum X_1^2X_2^2 + a_2\sum X_2^3 + b_2\sum X_2^4 = \sum YX_2^2
\end{aligned} \qquad (16.65)$$

Prin rezolvarea sistemului (16.65) se vor determina parametrii ecuaţiei regresiei multpile curbilinii: a, a_1, b_1, a_2, b_2.

Reprezentarea grafică a ecuaţiei regresiei multiple curbilinii se face printr-o suprafaţă în spaţiu ca şi în cazul legăturilor multiple liniare, dar se deosebeşte de acestea prin faptul că nu mai este plană (fig.16.5). In fig. 16.5 (A) se prezintă o legătură multiplă dintre Y, X_1 şi X_2 în cazul în care legăturile simple dintre variabile au forme diferite, astfel între Y şi X_1 există o legătură curbilinie, iar între Y şi X_2 o legătură liniară.

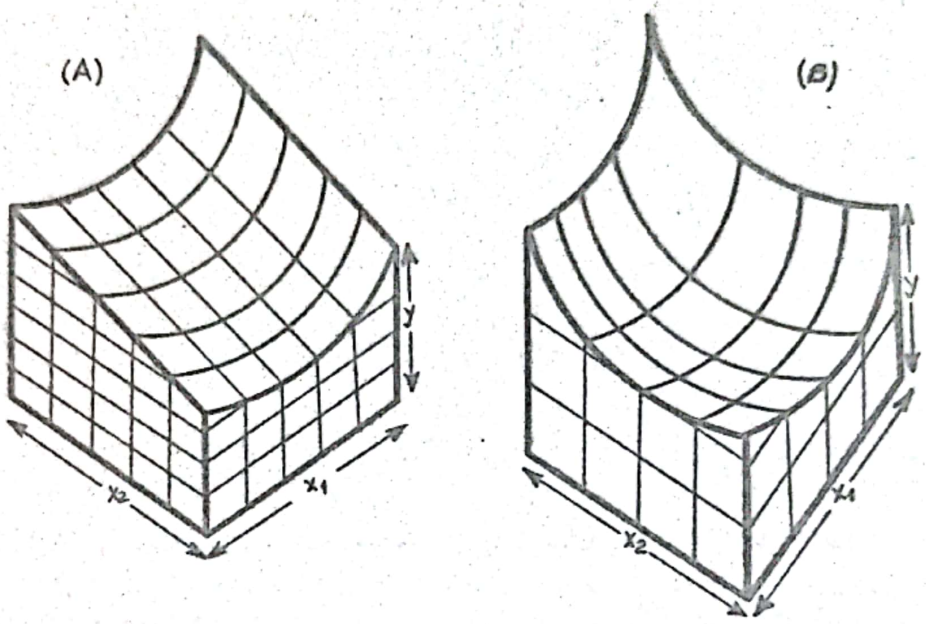


Fig. 16.5. LEGĂTURI MULTIPLE CURBILINII ÎNTRE TREI VARIABILE

Fig. 16.5 (B) prezintă forma suprafeţei de regresie în cazul în care regresia simplă dintre Y şi X_1, cît şi cea dintre Y şi X_2 sînt curbilinii.

16.3.2. Intensitatea corelaţiei multiple curbilinii

Ca şi în cazul corelaţiei multiple de formă liniară, intensitatea corelaţiei multiple curbilinii se măsoară cu ajutorul raportului de corelaţie:

$$\eta_{y.x_1x_2\ldots x_k} = \sqrt{1 - \frac{\sigma^2_{y.yx_1x_2\ldots x_k}}{\sigma^2_y}}$$

care prin substituirea celor două dispersii cu expresiile lor, devine:

$$\eta_{y.x_1x_2\ldots x_k} = \sqrt{1 - \frac{\sum(Y_j - Y_{x_1x_2\ldots x_k})^2}{\sum(Y_j - \bar{y})^2}} \qquad (16.66)$$

In cazul regresiei curbilinii de tipul polinomului de gradul doi, sumele de pătrate din formula (16.66) devin :

$$\sum(Y_j - Y_{x_1x_2\ldots x_k})^2 = \sum(Y_j - Y_{x_1x_2\ldots x_k})\,Y_j =$$

$$= \sum(Y_j - a - a_1X_1 - b_1X_1^2 - a_2X_2 - b_2X_2^2 - \ldots - a_iX_i - b_iX_i^2 - \ldots - a_kX_k - b_kX_k^3)Y_j =$$

$$= \sum Y_j^2 - a\sum Y_j - a_1\sum X_1Y_j - b_1\sum X_1^2Y_j - a_2\sum X_2Y_j - b_2\sum X_2^2Y_j -$$

$$- \ldots - a_i\sum X_iY_j - b_i\sum X_i^2Y_j - \ldots - a_k\sum X_kY_j - b_k\sum X_k^2Y_j \qquad (16.67)$$

Substituind pe (16.67) în (16.66) rezultă:

$$\eta_{y.x_1x_2\ldots x_k} = \sqrt{\frac{a\sum Y_j + a_1\sum X_1Y_j + b_1\sum X_1^2Y_j + a_2\sum X_2Y_j + b_2\sum X_2^2Y_j + \ldots + a_i\sum X_iY_j + b_i\sum X_i^2Y_j + \ldots + a_k\sum X_kY_j + b_k\sum X_k^2Y_j - \frac{(\sum Y_j)^2}{n}}{\sum Y_j^2 - \frac{(\sum Y_j)^2}{n}}} \qquad (16.68)$$

In particular, pentru regresiile curbilinii polinomiale din trei variabile, raportul de corelaţie (16.68) devine:

$$\eta_{y.x_1x_2} = \sqrt{\frac{a_1\sum Y_j + a_1\sum X_1Y_j + b_1\sum X_1^2Y_j + a_2\sum X_2Y_j + b_2\sum X_2^2Y_j - \frac{(\sum Y_j)^2}{n}}{\sum Y_j^2 - \frac{(\sum Y_j)^2}{n}}} \qquad (16.69)$$

In cazul a patru variabile raportul de corelaţie (16.68) va fi:

$$\eta_{y.x_1x_2x_3} = \sqrt{\frac{a_1\sum Y_j + a_1\sum X_1Y_j + b_1\sum X_1^2Y_j + a_2\sum X_2Y_j + b_2\sum X_2^2Y_j + a_3\sum X_3Y_j - b_3\sum X_3^2Y_j - \frac{(\sum Y_j)^2}{n}}{\sum Y_j^2 - \frac{(\sum Y_j)^2}{n}}} \qquad (16.70)$$

16.3.3. Determinarea curbilinie multiplă şi parţială

Ca şi în cazurile anterioare, coeficientul determinaţiei multiple curbilinii se obţine prin ridicarea la pătrat a raportului de corelaţie:

$$D_{y.x_1x_2\ldots x_k} = \eta^2_{y.x_1x_2\ldots x_k}$$

In particular, pentru cazul legăturilor curbilinii descrise de modelul polinomului de gradul doi, coeficientul determinaţiei multiple devine:

$$D_{y.x_1x_2\ldots x_k} = \frac{a\sum Y_j + a_1\sum X_1Y_j + b_1\sum X_1^2Y_j + a_2\sum X_2Y_j + b_2\sum X_2^2Y_j + \ldots + a_1\sum X_iY_j + b_1\sum X_i^2Y_j + \ldots + a_k\sum X_kY_j + b_k\sum X_k^2Y_j - \frac{(\sum Y_j)}{n}}{\sum Y_j^2 - \frac{(\sum Y_j)}{n}} \qquad (16.71)$$

Mărimea influenţei fiecăruia dintre cei k factori, care acţionează asupra fenomenului efect, se poate măsura cu ajutorul coeficienţilor de determinaţie parţială. Formulele lor de calcul se obţin prin substituirea parametrului a în (16.71) cu valoarea lui care rezultă din prima ecuaţie a sistemului de ecuaţii normale (16.62).

$$a = \bar{y} - a_1\bar{x}_1 - b_1\bar{x}_1^2 - a_2\bar{x}_2 - b_2\bar{x}_2^2 - \ldots - a_1\bar{x}_1 - b_1\bar{x}_1^2 - \ldots - a_k\bar{x}_k - b_k\bar{x}_k^2$$

Astfel:

$$D_{yx_1x_2\ldots x_k} =$$

$$= \frac{\frac{(\sum Y_j)^2}{n} - a_1\frac{\sum X_1\sum Y_j}{n} - b_1\frac{\sum X_1^2Y_j}{n} - a_2\frac{\sum X_2Y_j}{n} - b_2\frac{\sum X_2^2Y_j}{n} - \ldots - a_i\frac{\sum X_iY_j}{n} - b_i\frac{\sum X_i^2Y_j}{n} - \ldots - a_k\frac{\sum X_kY_j}{n} - b_k\frac{\sum X_k^2Y_j}{n} +}{}$$

$$+ \frac{+a_1\sum X_1Y_j + b_1\sum X_1^2Y_j + a_2\sum X_2Y_j + b_2\sum X_2^2Y_j + \ldots + a_i\sum X_iY_j + b_i\sum X_i^2Y_j + \ldots + a_k\sum X_kY_j + b_k\sum X_k^2Y_j - \frac{(\sum Y_j)^2}{n}}{\sum Y_j^2 - \frac{(\sum Y_j)^2}{n}} =$$

$$= \frac{a_1(n\sum X_1Y_j - \sum X_1\sum Y_j) + b_1(n\sum X_1^2Y_j - \sum X_1Y_j)}{n\sum Y_j^2 - (\sum Y_j)^2} + \frac{a_2(n\sum X_2Y_j - \sum X_2\sum Y_j) + b_2(n\sum X_2^2Y_j - \sum X_2^2\sum Y_j)}{n\sum Y_j^2 - (\sum Y_j)^2} + \ldots$$

$$\ldots \frac{a_i(n\sum X_iY_j - \sum X_iY_j) + b_i(n\sum X_i^2Y_j - \sum X_i^2\sum Y_j)}{n\sum Y_j^2 - (\sum Y_j)^2} + \ldots + \frac{a_k(n\sum X_kY_j - \sum X_k\sum Y_j) + b_k(n\sum X_k^2Y_j - \sum X_k^2\sum Y_k)}{n\sum Y_j^2 - (\sum Y_j)^2} = \tag{16.73}$$

$$= d_{yx_1x_2\ldots x_k} + d_{yx_2x_1x_3\ldots x_k} + \ldots + d_{yx_i\cdot x_1x_2\ldots x_k} + \ldots + d_{yx_k\cdot x_1x_2\ldots x_{k-1}}$$

Din relația (16.73) rezultă formulele coeficienților de determinație parțială, în cazul legăturii curbilinii luate în calcul:

$$d_{yx_1\cdot x_2x_3\ldots x_k} = \frac{a_1(n\sum X_1Y_j - \sum X_1\sum Y_j) + b_1\sum X_1^2Y_j - \sum X_1^2\sum Y_j)}{n\sum Y_j^2 - (\sum Y_j)^2} \tag{16.74}$$

$$d_{yx_2 \cdot x_1 x_3 \ldots x_k} = \frac{a_2(n\sum X_2 Y_j - \sum X_2 \sum Y_j) + b_2(n\sum X_2^2 Y_j - \sum X_2^2 \sum Y_j)}{n\sum Y_j^2 - (\sum Y_j)^2} \qquad (16.75)$$

.....

$$d_{yx_i \cdot x_1 x_2 \ldots x_k} = \frac{a_i(n\sum X_i Y_j - \sum X_i \sum Y_j) + b_i(n\sum X_i^2 Y_j - \sum X_i^2 \sum Y_j)}{n\sum Y_j^2 - (\sum Y_j)^2} \qquad (16.76$$

.....

$$d_{yx_k \cdot x_1 x_2 \ldots x_{k-1}} = \frac{a_k(n\sum X_k Y_j - \sum X_k \sum Y_j) + b_k(n\sum X_k^2 Y_j - \sum X_k^2 \sum Y_j)}{n\sum Y_j^2 - (\sum Y_j)^2} \qquad (16.77)$$

Deci

$$D_{y \cdot x_1 x_2 \ldots x_k} = d_{yx_1 \cdot x_2 x_3 \ldots x_k} + d_{yx_2 \cdot x_1 x_3 \ldots x_k} + \ldots d_{yx_i \cdot x_1 x_2 \ldots x_k} + \ldots + d_{yx_k \cdot x_1 \ldots x_{k-1}} \qquad (16.78)$$

16.3.4. Liniarizarea legăturilor multiple curbilinii:

Creşterea economică, în anumite condiţii, se poate estima, după cum se ştie, cu ajutorul modelului descris de o funcţie de tip Cobb-Douglas:

$$y = a\, x_1^{b_1}\, x_2^{b_2} \qquad (16.79)$$

Această funcţie se poate liniariza prin logaritmare şi devine:

$$\log Y_x = \log a + b_1 \log X_1 + b_2 \log X_2 \qquad (16.80)$$

Forma liniarizată se pretează cu uşurinţă la aplicarea metodei celor mai mici pătrate, astfel:

$$S = \sum(\log Y_j - \log Y_x)^2 = \text{minim}$$

sau

$$S = \sum(\log Y_j - \log a - b_1 \log X_1 - b_2 \log X_2)^2 = \text{minim}$$

Calculînd derivatele parţiale şi egalîndu-le cu zero, rezultă următorul sistem de ecuaţii normale:

$$n \log a + b_1 \sum \log X_1 + b_2 \sum \log X_2 = \sum \log Y_j \qquad (16.81)$$

$$\log a \sum \log X_1 + b_1 \sum (\log X_1)^2 + b_2 \sum \log X_2 \log X_1 = \sum \log Y_j \log X_1$$

$$\log a \sum \log X_2 + b_1 \sum \log X_1 \log X_2 + b_2 \sum \log X_2^2 = \sum \log Y_j \log X_2$$

Sistemul (16.81) se poate rezolva prin metodele clasice de rezolvare a sistemelor de ecuaţii normale.

In cazul regresiei multiple curbilinii descrise de modelul (16.4) liniarizarea se obţine tot prin logaritmare, astfel:

$$\log Y_x = \log a + b_1 \log X_1 + b_2 \log X_2 + b_3 \log X_3 \qquad (16.82)$$

Parametrii modelului (16.82) se pot determina, de asemenea, cu ajutorul metodei celor mai mici pătrate.

După liniarizare, indicatorii intensităţii corelaţiei se pot determina în mod asemănător cu cazul legăturilor liniare multiple.

17. CORELATIA NEPARAMETRICA

Metodele de studiere a corelaţiei dintre fenomene, prezentate pînă acum, se aplică în practică numai în cazul variabilelor distribuite normal sau apropiate de normalitate. Aceste metode poartă denumirea de metode parametrice. Dacă distribuţia variabilelor corelate nu este normală, pentru măsurarea intensităţii dintre variabile, se pot utiliza metode neparametrice. Metodele neparametrice au fost elaborate pentru prima dată de către C. Spearman şi au fost dezvoltate de către K. Pearson, M. Kendall şi alţi statisticieni.

In utilizarea metodelor neparametrice se face abstracţie de tipul distribuţiilor variabilelor corelate, fapt care le măreşte substanţial cadrul de aplicabilitate. Metodele neparametrice nu operează cu valorile reale ale caracteristicilor, ci cu rangurile lor.

In acest scop, unităţile statistice se ordonează în ordine crescătoare sau descrescătoare a variabilei factoriale şi apoi se atribuie fiecărei variante cîte un rang. Deoarece nu operează direct cu nivelurile caracteristicilor, aceste metode sînt aplicabile şi în cazul corelaţiei dintre caracteristici calitative.

Dacă între variabilele corelate există o legătură directă, atunci unităţile statistice care s-a clasat pe primele locuri în raport cu caracteristica factorială se vor clasa tot pe primele locuri şi în raport cu caracteristica rezultativă, în timp ce în cazul unei legături inverse, rangurilor mici ale variabilei factoriale le vor corespunde rangurile mari ale variabilei rezultative.

In primul caz există o concordanţă a rangurilor celor două variabile, iar în cazul al doilea o discordanţă de ranguri. Concordanţa dintre ranguri are loc în cazul unor legături directe, iar discordanţa în cazul legăturilor inverse.

Astfel, între gradul de industrializare a judeţelor României şi gradul lor de urbanizare există o legătură statistică di-

rectă, în sensul că pe măsură ce creşte gradul de dezvoltare industrială a judeţelor va creşte şi forţa de muncă necesară pentru industrie, fapt care va determina o creştere a gradului de urbanizare. Aceasta înseamnă că o comparaţie a rangurilor judeţelor după cele două caracteristici va da posibilitatea să se măsoare intensitatea legăturii statistice dintre cele două fenomene.

Dintre metodele neparametrice de studiere a corelaţiei dintre fenomene vom prezenta, în continuare, metoda lui Spearman, metoda lui Kendall, metoda lui Fechner.

2.1. Coeficientul de corelaţie a rangurilor al lui Spearman

In cazul unei legături funcţionale directe dintre variabile ar trebui să existe o concordanţă deplină între rangurile celor două variabile corelate şi o discordanţă totală în cazul legăturilor funcţionale inverse.

Pe baza acestei ipoteze, C. Spearman[1] a stabilit un coeficient de corelaţie a rnagurilor, dezvoltat şi caracterizat sub multiple aspecte de către K. Pearson[2] şi cunoscut sub următoarea formulă:

$$\theta = 1 - \frac{6 \sum d_i^2}{n^3 - n} \qquad (17.1)$$

în care: d_i - reprezintă diferenţele dintre rangurile celor două variabile;

n - numărul perechilor de valori corelate.

Coeficientul lui Spearman ca şi coeficientul r_{xy} al lui Pearson variază între limitele -1 şi +1 cu aceeaşi semnificaţie:

$$-1 \leq \theta \leq +1 \qquad (17.2)$$

Formula coeficientului de corelaţie a rangurilor a lui Spearman se poate deduce şi pe baza coeficientului de corelaţie al lui Pearson, notînd rangurile şi simbolurile utilizate pentru variantele variabilelor corelate:

---

1) Spearman C. : "A foot-rule for measuring correlation", British Journal pf Psychology, vol.2, p. 89.

2) Pearson K. : "On further methods of determining correlation", Drapers'Company Research Memoires, Biometric Series IV.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n\,\sigma_x\,\sigma_y} \qquad (17.3)$$

Considerînd că rangurile variabilelor X şi Y sînt constituite din primele n numere întregi şi pozitive, nivelul mediu al celor două variabile va fi dat de expresia:

$$\bar{x} = \bar{y} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2} \qquad (17.4)$$

iar abaterile medii pătratice se vor calcula astfel:

$$\sigma_x = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2} \quad \text{şi} \quad \sigma_y = \sqrt{\frac{\sum y_i^2}{n} - \bar{y}^2} \qquad (17.5)$$

Suma pătratelor primelor n numere naturale este dată de relaţia:

$$1^2+2^2+3^2+\ldots+n^2 = \frac{n(n+1)(2n+1)}{6} = \sum x_i^2 = \sum y_i^2 \qquad (17.6)$$

$$\sigma_x = \sigma_y = \sqrt{\frac{\frac{n(n+1)(2n+1)}{6}}{n} - \left(\frac{n+1}{2}\right)^2} =$$

$$= \sqrt{\frac{n(n+1)(2n+1)}{6n} - \frac{(n+1)^2}{4}} = \sqrt{\frac{2(n+1)(2n+1)-3(n+1)^2}{12}} = \sqrt{\frac{n^2-1}{12}}$$

$$n\,\sigma_x\,\sigma_y = n \cdot \frac{n^2-1}{12} \qquad (17.7)$$

Notăm diferenţa dintre rangurile celor două variabile cu $d_i = x_i - y_i$ sau adunînd şi scăzînd aceeaşi cantitate $\bar{x} = \bar{y}$ rezultă: $d_i = (x_i - \bar{x}) - (y_i - \bar{y})$.

Ridicînd la pătrat ambii membri ai acestei relaţii şi însumînd, se obţine:

$$\sum d_i^2 = \sum[(x_i-\bar{x})-(y_i-\bar{y})]^2 = \sum[(x_i-\bar{x})^2-2(x_i-\bar{x})(y_i-\bar{y})+(y_i-\bar{y})]^2 =$$

$$= \sum(x_i-\bar{x})^2-2\sum(x_i-\bar{x})(y_i-\bar{y})+\sum(y-\bar{y})^2$$

dar

$$\sum(x_i - \bar{x})^2 = \sum(y_i - \bar{y})^2 = \frac{n(n^2 - 1)}{12}$$

deci

$$\sum d_i^2 = \frac{2n\,(n^2 - 1)}{12} - 2\sum(x_i - \bar{x})(y_i - \bar{y})$$

de unde

$$\sum(x_i - \bar{x})(y_i - \bar{y}) = \frac{n(n^2 - 1)}{12} - \frac{6\sum d_i^2}{12} \qquad (17.8)$$

Substituind pe (17.7) şi (17.8) în formula coeficientului de corelaţie Pearson, se obţine:

$$r = \frac{\frac{n(n^2-1)}{12} - \frac{6\sum d_i^2}{12}}{\frac{n(n^2-1)}{12}} = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = \Theta$$

In cazul concordanţei depline a rangurilor, toate diferenţele $d = x_i - y_i$ sînt nule şi deci $\Theta = +1$; iar în cazul a două serii de ranguri total opuse $\Theta = -1$.

Pentru exemplificarea calculului şi analizei coeficientului de corelaţie a rangurilor al lui Spearman, să considerăm corelaţia dintre dezvoltarea industrială a judeţelor şi gradul lor de urbanizare (tabelul 17.1).

La stabilirea rangurilor celor două variabile corelate se poate întîmpla ca două sau mai multe variante ale unei variabile să aibă aceeaşi valoare, fapt care nu ne permite să atribuim fiecărei variante un rang unic şi bine determinat. In exemplul nostru, judeţele Caraş-Severin şi Maramureş au acelaşi grad de urbanizare (variabila Y) şi anume 47,4%. Pentru soluţionarea acestei probleme se poate proceda în două moduri şi anume, presupunem că variantele sînt 5,6,7,7,7,8,lo,12.

a) Se poate atribui fiecăreia dintre variantele egale ale unei variabile un rang egal cu media rangurilor pe care aceste variante le ocupă, astfel: 1,2,4,4,4,6,7,8;

b) Se poate atribui fiecăreia dintre variantele egale rangul pe care îl ocupă prima dintre aceste variante, astfel: 1,2,3,3, 3,6,7,8.

Tabelul 17.1. Producţia globală industrială şi populaţia urbană pe judeţe în anul 1975

Judeţul	Ponderea populaţiei urbane (Y)	Producţia globală industrială (milioane lei (X)	Rangul judeţului după		d_1	d_1^2
			X	Y		
A	1	2	3	4	5	6
Alba	41,5	8106	23	14	9	81
Arad	46,9	11585	17	12	5	25
Argeş	34,0	21937	4	23	19	361
Bacău	39,1	16336	12	16	4	16
Bihor	37,8	14036	14	18	4	16
Bistriţa Năsăud	19,6	1808	39	38	1	1
Botoşani	21,7	3388	36	35	1	1
Braşov	65,3	38295	1	2	1	1
Brăila	48,3	11198	20	8	12	144
Buzău	21,5	9059	22	36	14	196
Caraş Severin	47,4	11478	18	10	8	64
Cluj	54,0	20334	8	5	3	9
Constanţa	59,2	12842	15	3	13	169
Covasna	42,9	3494	35	13	22	484
Dîmboviţa	33,7	11372	19	24	5	25
Dolj	37,7	18073	10	19	9	81
Galaţi	49,0	25425	3	7	4	16
Gorj	29,3	5802	30	27	3	9
Harghita	37,1	6579	28	20	8	64
Hunedoara	70,9	21503	5	1	4	16
Ialomiţa	28,2	5272	32	28	4	16
Iaşi	38,6	19113	9	17	8	64
Ilfov	13,6	7433	26	39	13	169
Maramureş	47,4	7952	24	10	14	196
Mehedinţi	36,2	6447	29	22	7	49
Mureş	41,4	17171	11	15	4	16
Neamţ	32,2	15933	13	25	12	144
Olt	22,2	13456	15	34	19	361
Prahova	51,8	32247	2	6	4	16
Satu Mare	36,0	7688	25	21	4	16
Sălaj	20,6	2082	38	37	1	1
Sibiu	57,8	21287	6	4	2	4
Suceava	27,7	9928	21	29	8	64
Teleorman	26,3	6970	27	30	3	9
Timiş	47,9	20659	7	9	2	4
Tulcea	30,5	2675	37	26	11	121
Vaslui	25,2	4907	33	31	2	4
Vîlcea	24,3	5638	31	32	1	1
Vrancea	24,0	4223	34	33	1	1
TOTAL	-	-	-	-	-	3035

Sursa: Anuarul statistic al R.S.R. 1976, DCS, Bucureşti, p. 12-13, 94.

Remarcăm faptul că, în ambele cazuri, la trecerea de la grupa de variante egale la următoarea unitate statistică se consideră rangul primei unități ce urmează ca și cum toate elementele precedente ar fi fost negale. Cînd numărul perechilor de valori corelate este suficient de mare nu are importanță ce metodă se utilizează, dar în cazul unui număr redus de perechi de valori corelate, coeficientul de corelație al lui Spearman va avea valori diferite.

In tabelul 17.1, după ce au fost stabilite rangurile (și soluționată problema celor două județe cu același nivel al variabilei Y, după metoda a doua), s-au determinat, în coloana 5, diferențele dintre ranguri și, în coloana 6, pătratele acestor diferențe.

Se remarcă faptul că este indiferent dacă s-au ordonat sau nu județele (unitățile statistice în general) după variația variabilei independente, deoarece și într-un caz și în celălalt diferențele dintre ranguri d_i, pentru unitatea statistică i , rămîn neschimbate.

Aplicînd formula coeficientului lui Spearman (17.1), rezultă:

$$\theta = 1 - \frac{6\sum d_i^2}{n^3-n} = 1 - \frac{6.3035}{39^3-39} = 0,69$$

Deoarece $\theta = 0,69 > 0$ înseamnă că între gradul de industrializare a județelor și gradul de urbanizare există o legătură directă și relativ intensă.

17.2. Coeficientul de corelație a rangurilor al lui Kendall

Un alt coeficient de corelație, bazat pe rangurile variabilelor corelate, a fost propus de statisticianul engels M.G. Kendall:

$$\tau = \frac{S}{\frac{1}{2} n (n-1)} \qquad (17.9)$$

în care $S = P - Q$, iar $P = \sum p_i$ și $Q = \sum q_i$

Cu p_i s-a notat numărul de ranguri superioare ale variabilei dependente rangului i care există pe coloana respectivă după fiecare rang; iar q_i reprezintă numărul de ranguri inferioare ale

variabilei dependente, care există după fiecare rang al acesteia.

Coeficientul lui Kendall variază tot între -1 şi +1. Astfel, dacă între rangurile celor două variabile există o concordanţă deplină, atunci $Q = 0$, iar $\tau = 1$, deci între rânguri există o concordanţă totală, atunci $P = 0$ şi $\tau = -1$.

Pentru exemplificarea calculului şi analizei coeficientului τ să considerăm legătura dintre dezvoltarea industrială şi gradul de urbanizare a judeţelor României. Se înscriu judeţele în ordine crescătoare a rangurilor variabilei independente (producţia globală industrială) şi pe coloana următoare se trec rangurile corespunzătoare ale variabilei dependente (gradul de urbanizare) ca în tabelul 17.2. Pe coloanele următoare se înscriu valorile lui p_i şi apoi cele ale lui q_i.

Aplicînd formula (17.9) rezultă:

$$S = P - Q = 565 - 176 = 389$$

$$\tau = \frac{389}{\frac{1}{2}\,39(39-1)} = \frac{389}{741} = 0{,}52$$

$\tau = 0{,}52$ înseamnă că legătura dintre dezvoltarea industrială şi gradul de urbanizare a judeţelor este directă şi de intensitate medie. Remarcăm că τ este întotdeauna mai mic ca valoare decît θ.

In cazul unor valori suficient de mari ale lui n, profesorul Kendall a arătat că între coeficienţii de corelaţie a rangurilor θ şi τ are loc următoarea relaţie aproximativă (cu excepţia cazurilor în care cei doi coeficienţi se apropie ca valoare de unitate):

$$\theta = \frac{3\tau}{2} \qquad (17.10)$$

In exemplul nostru $\theta = 0{,}69$ şi $\tau = 0{,}52$, deci $\theta \simeq \frac{3 \cdot 0{,}52}{2} = 0{,}78$.

Faţă de θ, calculul coeficientului de corelaţie τ al lui Kendall este mai dificil dar prezintă o serie de avantaje faţă de primul[x].

x G.U. Yule, M.G. Kendall, "Introducere în teoria statisticii", trad. din engleză după ed. a XIV-a, Ed. ştiinţifică, Bucureşti, 1969, p.281.

Tabelul 17.2. Producţia globală industrială şi ponderea populaţiei urbane pe judeţe. Calculul coeficientului de corelaţie al lui Kendall

J u d e ţ u l	Rangul judeţului după		p_i	q_i
	X	Y		
Braşov	1	2	37	1
Prahova	2	6	33	4
Galaţi	3	7	32	4
Argeş	4	23	16	19
Hunedoara	5	1	34	0
Sibiu	6	4	32	1
Timiş	7	9	29	3
Cluj	8	5	30	1
Iaşi	9	17	21	9
Dolj	10	19	19	10
Mureş	11	15	21	7
Bacău	12	16	20	7
Neamţ	13	25	14	12
Bihor	14	18	18	7
Olt	15	34	5	19
Constanţa	16	3	23	0
Arad	17	12	19	3
Caraş Severin	18	10	20	1
Dîmboviţa	19	24	13	7
Brăila	20	8	19	0
Suceava	21	29	9	9
Buzău	22	35	3	14
Alba	23	14	14	2
Maramureş	24	10	15	0
Satu Mare	25	21	12	2
Ilfov	26	39	0	13
Teleorman	27	30	6	6
Harghita	28	20	10	1
Mehedinţi	29	22	9	1
Gorj	30	27	7	2
Vîlcea	31	32	4	4
Ialomiţa	32	28	5	2
Vaslui	33	31	4	2
Vrancea	34	33	3	2
Covasna	35	13	4	0
Botoşani	36	35	2	1
Tulcea	37	25	2	0
Sălaj	38	37	1	0
Bistriţa Năsăud	39	38	-	-
T O T A L	-	-	565	176

a) Se poate verifica, prin testele obişnuite în statistică, semnificaţia lui τ în timp ce pentru θ aceasta se poate face numai într-un caz particular, cînd n este suficient de mare;

b) coeficientul lui Kendall poate fi utilizat şi pentru studiul corelaţiei parţiale a rangurilor.

Pentru compararea intensităţii legăturii dintre diferite fenomene, este necesar să se utilizeze acelaşi indicator al intensităţii corelaţiei. Este cu totul greşit să se utilizeze pentru comparaţie indicatori diferiţi de corelaţie, de exemplu, coeficientul lui Pearson şi un coeficient de corelaţia a rangurilor, deoarece între aceşti indicatori există importante deosebiri metodologice. Mai mult decît atît, coeficienţii de corelaţie simplă, pe care i-am tratat în acest capitol, vor avea valori diferite chiar dacă se calculează pentru aceleaşi perechi de valori corelate.

17.3. Coeficientul de concordanţă al lui Fechner

Statisticianul german G. Fechner[x] a stabilit la sfîrşitul secolului al XIX-lea un indicator al intensităţii corelaţiei calculat pe baza concordanţelor şi discordanţelor diferenţelor dintre termenii consecutivi ai fiecărei dintre cele două variabile corelate, fie cu ajutorul concordanţelor şi respectuv discordanţelor abaterilor calculate faţă de mediile varibilelor corelate.

Coeficientul de concordanţă propus de Fechner, cunoscut sub denumirea <u>coeficient simplu de covariaţie diferenţială</u>, se calculează cu ajutorul formulei:

$$k = \frac{c - d}{n} \qquad (17.11)$$

în care c şi d reprezintă numărul de concordanţe, respectiv de discordanţe, de semne ale abaterilor:

$$\Delta x_i = x_i - x_{i-1} \quad \text{şi} \quad \Delta y_i = y_i - y_{i-1}$$

sau

$$\Delta x_i = x_i - \bar{x} \quad \text{şi} \quad \Delta y_i = y_i - \bar{y}$$

şi n = numărul perechilor de valori corelate.

x) G. Fechner: Kollektivmasslehre, Leipzig, 1897.

Dacă una sau mai multe dintre diferenţele Δ_{x_i} sau Δ_{y_i} este nulă, atunci, pentru perechea respectivă x_i, y_i a celor două variabile nu se consideră nici concordanţă şi nici discordanţă, ci este exclusă din calcul.

Din formula (17.11) rezultă că acest indicator variază între -1 şi +1 .

Astfel, dacă între Δ_{x_i} şi Δ_{y_i} există numai concordanţe de semne, înseamnă că $c = n$ şi $d = 0$ iar $k = 1$; dacă între Δ_{x_i} şi Δ_{y_i} există numai discordanţe de semne, atunci $d = n$ şi $c = 0$, iar coeficientul de concordanţă $k = -1$; k este nul în cazul în care numărul concordanţelor este egal cu numărul discordanţelor $c = d$ şi $k = 0$.

Coeficientul k de concordanţă prezintă dezavantajul că nu ia în calcul mărimea abaterilor Δ_{x_i} şi Δ_{y_i}, ci ţine seama numai de semnele acestor abateri. Pentru a se înlătura această deficienţă se calculează aşa-numitul <u>coeficient ponderat de concordanţă</u> după formula:

$$K = \frac{C - D}{C + D} \qquad (17.12)$$

unde C - reprezintă suma produselor $\Delta_{x_i} \Delta_{y_i}$ pozitive ,

D - reprezintă valoarea absolută a sumei produselor $\Delta_{x_i} \Delta_{y_i}$ negative.

La fel ca şi k, coeficientul K variază între -1 şi +1 şi are aceeaşi semnificaţie. Dacă $D = 0$ atunci $K = 1$ şi dacă $C = 0$ atunci $K = -1$, iar cînd $C = D$ coeficientul ponderat de concordanţă este nul.

Atît coeficientul simplu de concordanţă cît şi cel ponderat au dezavantajul că ei pot lua valorile ± 1, chiar atunci cînd legătura dintre variabile nu este funcţională, adică diferenţele Δ_{x_i} şi Δ_{y_i} nu sînt proporţionale. Acest neajuns se înlătură dacă se utilizează o altă variantă a coeficientului ponderat de concordanţă, mai apropiată ca structură de coeficientul de corelaţie al lui Bravais-Pearson:

$$K_c = \frac{\sum \Delta_{x_i} \Delta_{y_i}}{\sqrt{\sum (\Delta_{x_i})^2 \sum (\Delta_{y_i})^2}} \qquad (17.13)$$

Tabelul 17.3. Producţia globală industrială şi ponderea populaţiei urbane pe judeţele R.S.R. în anul 1976. Calculul coeficienţilor de concordanţă

Judeţul	Valoarea prod. glob.ind. (mld.lei)	Ponderea populaţiei urbane la (%) 1-ol-1977	Δ_{x_i}	Δ_{y_i}	c sau d	$(\Delta_{x_i})^2$	$(\Delta_{y_i})^2$	$\Delta_{x_i}\Delta_{y_i}$
A	1	2	3	4	5	6	7	8
Alba	9,2	44,5	-	-	-	-	-	-
Arad	13,2	50,6	4,0	-6,1	c	16,00	37,21	24,40
Argeş	26,8	38,4	-13,6	12,2	d	184,96	148,84	-165,92
Bacău	18,1	42,1	8,7	-3,7	d	75,69	13,69	-32,19
Bihor	15,5	40,1	2,6	2,0	c	6,76	4,00	5,20
Bistriţa N.	2,2	23,8	13,3	16,3	c	176,89	265,69	216,79
Botoşani	4,1	26,1	- 1,9	-2,3	c	3,61	5,29	4,37
Braşov	43,8	70,0	-39,7	-43,9	c	1576,09	1927,21	1742,83
Brăila	12,6	54,4	-31,2	15,6	c	973,44	243,36	486,72
Buzău	10,9	24,9	1,7	29,5	c	2,89	870,25	50,15
Caraş Sev.	13,0	49,9	- 2,1	25,0	d	4,41	625,00	-52,50
Cluj	22,2	58,3	- 9,2	- 8,4	c	84,64	70,56	77,28
Constanţa	14,4	67,1	7,8	- 8,8	d	60,84	77,44	-68,64
Covasna	3,9	46,6	10,5	20,5	c	110,25	420,25	215,25
Dîmboviţa	13,6	37,0	- 9,7	9,6	d	94,09	92,16	-93,12
Dolj	19,9	41,0	- 6,3	- 4,0	c	39,69	16,00	25,20
Galaţi	29,6	54,2	- 9,7	-13,2	c	94,09	174,24	128,04
Gorj	6,4	32,4	23,2	21,8	c	538,24	475,24	505,76
Harghita	7,4	40,2	- 1,0	- 7,8	c	1,00	60,84	7,80
Hunedoara	23,0	71,8	-15,6	-31,6	c	243,36	998,56	492,96
Ialomiţa	5,9	33,5	17,1	38,3	c	292,41	1466,89	654,93
Iaşi	22,6	45,6	-16,7	-12,1	c	278,89	146,41	202,07
Ilfov	8,1	15,1	14,5	30,5	c	210,25	930,25	442,25
Maramureş	8,9	48,7	- 0,8	-33,6	c	0,64	1128,96	26,88
Mehedinţi	6,6	38,7	2,3	10,0	c	5,29	100,00	23,00
Mureş	19,6	45,2	-13,0	- 6,5	c	169,00	42,25	84,50
Neamţ	17,7	34,6	1,9	10,6	c	3,61	112,36	20,14
Olt	15,0	25,3	2,7	9,3	c	7,29	86,49	25,11
Prahova	35,6	53,9	-20,6	-28,6	c	424,36	817,96	589,16
Satu Mare	8,7	39,3	26,9	14,6	c	723,61	213,16	392,74
Sălaj	2,9	24,2	5,8	15,1	c	33,64	228,01	87,58
Sibiu	23,4	60,7	-20,5	-36,5	c	420,25	1332,25	748,25
Suceava	10,9	30,3	12,5	30,4	c	156,25	924,16	380,00
Teleorman	7,8	28,8	3,1	1,5	c	9,61	2,25	4,65
Timiş	23,3	53,1	-15,5	-24,3	c	240,25	590,49	376,65
Tulcea	3,1	35,8	20,2	17,3	c	408,04	299,29	349,46
Vaslui	6,1	28,7	- 3,0	7,1	d	9,00	50,41	-21,30
Vîlcea	7,6	28,3	- 1,5	0,4	d	2,25	0,16	- 6,00
Vrancea	4,8	27,6	2,8	0,7	c	7,84	0,49	1,96
T O T A L	-	-	-	-	-	7689,42	14997,51	+5913,72 - 439,67

Sursa: Rotunjit după Anuarul statistic al Republicii Socialiste România 1977, D.C.S., Bucureşti, p.49 şi p.130.

. Acest coeficient ia valoarea ± 1 numai în cazul în care Δ_{y_i} variază direct proporţional cu Δ_{x_i} ;

Pentru ilustrarea metodei de calcul şi analiză a acestor indicatori, să considerăm legătura dintre nivelul de dezvoltare industrială şi gradul de urbanizare a judeţelor din România în anul 1976. (Tabelul 17.3).

Coeficientul simplu de concordanţă:

$$k = \frac{a - d}{n} = \frac{31 - 7}{38} = \frac{24}{38} = 0,63$$

Coeficientul ponderat de concordanţă:

$$K = \frac{C - D}{C + D} = \frac{5913,73 - 439,67}{5913,73 + 439,67} = \frac{5474,06}{6353,40} = 0,86$$

Coeficientul pearsonian de concordanţă:

$$K_c = \frac{\sum \Delta_{x_i} \Delta_{y_i}}{\sqrt{\sum (\Delta_{x_i})^2 \sum (\Delta_{y_i})^2}} = \frac{5913,73 - 439,67}{\sqrt{7689,42 . 14997,51}} =$$

$$= \frac{5474,06}{10738,81} = 0,5097$$

Toţi aceşti trei coeficienţi de concordanţă arată că între nivelul de dezvoltare industrială şi gradul de urbanizare a judeţelor din România există o legătură directă, deoarece toţi coeficienţii sînt pozitivi. Valorile acestor coeficienţi sînt diferite datorită diferenţelor metodologiei dintre ei. Desigur, că între ei vom acorda mai multă încredere coeficientului pearsonian de concordanţă, care prezintă cele mai puţine deficienţe metodologice.

18. ASOCIATIA INTRE CARACTERISTICI CALITATIVE

18.1 Noţiunea de asociaţie

Indicatorii parametrici de corelaţie nu pot fi utilizaţi decît pentru caracterizarea legăturii dintre două sau mai multe variabile (caracteristici) exprimate numeric. Dacă variabilele a căror legătură se studiază, sînt de esenţă calitativă şi nu pot fi exprimate numeric (culori, categorii sociale, profesii, ocupaţii, opinii etc.), atunci legătura dintre ele se studiază cu ajutorul metodei asociaţiei.

Să considerăm două caracteristici (X) şi (Y), care sînt susceptibile de k şi respectiv h forme, astfel:

(X) $x_1, x_2, \ldots, x_i, \ldots, x_k$

(Y) $y_1, y_2, \ldots, y_j, \ldots, y_h$

Notăm: n_{ij} - numărul unităţilor care posedă simultan nivelul x_i şi y_i al caracteristicilor X şi Y ;

$n_{i\cdot}$ - numărul celor care posedă varianta x_i a variabilei X ;

$n_{\cdot j}$ - numărul celor care posedă varianta y_j a variabilei Y ;

n - numărul total al unităţilor.

Pe baza acestor notaţii se poate întocmi un tabel, în care unităţile să fie grupate concomitent cu ambele caracteristici studiate. Acest tabel poartă denumirea de <u>tabel de contingenţă</u> sau <u>tabel de asociaţie</u> (notat tabel kxh).

Tabelul 18.1. Schema unui tabel de contingenţă (kxh)

Y \ X	x_1	x_2	.....	x_i	....	x_k	TOTAL
y_1	n_{11}	n_{21}	.....	n_{i1}	....	n_{k1}	$n_{\cdot 1}$
y_2	n_{12}	n_{22}	.....	n_{i2}	....	n_{k2}	$n_{\cdot 2}$
⋮	⋮	⋮		⋮		⋮	⋮
y_j	n_{1j}	n_{2j}	.....	n_{ij}	....	n_{kj}	$n_{\cdot j}$
⋮	⋮	⋮		⋮		⋮	⋮
y_h	n_{1h}	n_{2h}	.....	n_{ih}	....	n_{kh}	$n_{\cdot h}$
TOTAL	$n_{1\cdot}$	$n_{2\cdot}$	.....	$n_{i\cdot}$	....	$n_{k\cdot}$	$n_{\cdot\cdot}$

Un caz mai simplu este acela în care fiecare dintre cele două caracteristici ale unităţilor cercetate este susceptibilă de două modalităţi. Tabelul de asociaţie în acest caz va avea următoarea formă (tabel notat 2 x 2):

Tabelul 18.2. Tabel de asociaţie 2 x 2

Y \ X	x_1	x_2	TOTAL
y_1	a	b	m_1
y_2	c	d	m_2
TOTAL	n_1	n_2	n

In general, studiul asociaţiei dintre fenomene se face în două direcţii principale :

- studiul existenţei asociaţiei;
- măsurarea intensităţii asociaţiei.

18.2. Existenţa asociaţiei

O caracteristică (X) este independentă de (Y) dacă proporţia unităţilor care sînt (x_i) şi (y_j) printre toate cele care sînt y_j este aceeaşi ca şi proporţia celor care sînt x_i faţă de ce-

lectivitatea totală (tabelul kxh) pentru orice valoare a lui i şi j, astfel:

$$\frac{n_{ij}}{n_{\cdot j}} = \frac{n_{i\cdot}}{n} \qquad \begin{matrix} i = 1,2, \ldots, k \\ j = 1,2, \ldots, h \end{matrix} \qquad (18.1)$$

de unde

$$n_{ij} = \frac{n_{i\cdot}\, n_{\cdot j}}{n} \qquad (18.2)$$

Această formulă sintetică ne arată că dacă variabila (X) este independentă de (Y), atunci şi (Y) este independentă de (X). Deci, dacă (X) şi (Y) sînt independente, atunci frecvenţa relativă din fiecare căsuţă i,j trebuie să fie egală cu produsul frecvenţelor relative marginale, adică:

$$\frac{n'_{ij}}{n} = \frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot j}}{n} \qquad (18.3)$$

de unde

$$n'_{ij} = \frac{n_{i\cdot}\, n_{\cdot j}}{n} \qquad (18.4)$$

Abaterile dintre frecvenţele empirice şi frecvenţele teoretice $n_{ij} - n'_{ij}$ măsoară dezacordul dintre caracteristicile studiate şi ipoteza de independenţă.

In cazul unui tabel de contingenţă 2 x 2, dacă notăm cu a,b,c,d frecvenţele empirice şi cu $\alpha, \beta, \gamma, \delta$ frecvenţele teoretice corespunzătoare, se obţine tabelul 18.3.

Sînt două metode mai des utilizate în statistică pentru studierea existenţei asociaţiei:

Metoda 1. Se verifică ipoteza de independenţă, adică dacă $p_1 = \frac{a}{n_1}$ este semnificativ diferit de $p = \frac{m_1}{n}$ sau dacă diferenţa $|a - \alpha|$ este semnific ativ diferită de zero, în care $\alpha = \frac{n_1 m_1}{n}$.

Se poate lua oricare dintre cele patru diferenţe între frecvenţele teoretice şi cele empirice pentru a caracteriza dezacordul dintre ele, deoarece:

$$|a - \alpha| = |b - \beta| = |c - \gamma| = |d - \delta| \qquad (18.5)$$

Tabelul 18.3. Tabel de asociaţie 2 x 2 cu frecvenţe teoretice şi empirice

X / Y	x_1	x_2	Total
y_1	$a^{(\alpha)}$	$b^{(\beta)}$	m_1
y_2	$c^{(\gamma)}$	$d^{(\partial)}$	m_2
Total	n_1	n_2	n

Pentru verificarea existenţei asociaţiei între două variabile, prezentate într-un tabel de asociaţie 2 x 2, se poate utiliza testul t :

$$t = \frac{a - \alpha}{\sigma} \qquad (18.6)$$

unde abaterea standard "σ" a diferenţei $a - \alpha$ este definită de relaţia:

$$\sigma^2 = n_1 p\, q\left(1 - \frac{n_1}{n}\right) = n_1 \frac{m_1}{n}\left(1 - \frac{n_1}{n}\right)\left(1 - \frac{n_1}{n}\right) = \frac{n_1 n_2 m_1 m_2}{n^3} \qquad (18.7)$$

Se determină pe baza tabelelor $P(t) = \frac{2}{\sqrt{2\pi}} \int_t^\infty e^{\frac{-t^2}{2}} dt$ (18.8)

în funcţie de care se acceptă sau se respinge ipoteza.

Metoda 2. Se caracterizează dezacordul dintre cele două feluri de frecvenţe (teoretice şi empirice) cu ajutorul criteriului χ^2 :

$$\chi^2 = \frac{(a-\alpha)^2}{\alpha} + \frac{(b-\beta)^2}{\beta} + \frac{(c-\gamma)^2}{\gamma} + \frac{(d-\partial)^2}{\partial} \qquad (18.9)$$

Se poate demonstra că $t^2 = \chi^2$, unde χ^2 are un singur grad de libertate. Pe baza acestei identităţi expresia lui χ^2 devine:

$$\chi^2 = \frac{(a-\alpha)^2}{\sigma^2} = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(a+c)(d+b)(d+c)} = \frac{n(ad-bc)^2}{n_1 n_2 m_1 m_2} \qquad (18.10)$$

Relaţia (18.10) arată că χ^2 este egal cu zero atunci cînd produsul de pe diagonală ad este egal cu produsul bc . Acest fapt este important, deoarece el ne va indica dacă este sau nu necesar să comntinuăm testul de semnificaţie.Astfel, dacă cele două produse sînt egale sau apropiate ca mărime, atunci χ^2 va fi prea mic pentru a valida ipoteza nulă. Produsele de pe diagonalele tabelului pot fi folosite, de asemenea, pentru a determina direcţia legăturii de asociaţie în sensul că produsul mai mare ne va indica concentrarea mai mare a valorilor şi deci direcţia legăturii.

Să considerăm pentru exemplificare distribuţia elevilor înscrişi în anul 1976/1977 după categoria liceului şi pe sexe, la recensămîntul din 5 ianuarie 1977 şi să verificăm dacă există vreo legătură între aceste două caracteristici.

Tabelul 18.4. Distribuţia elevilor înscrişi în anul şcolar 1976/1977 după felul liceului şi pe sexe (mii)

Licee Sexul	De specialitate	Real-umaniste şi de artă	Total
Masculin	371	171	542
Feminin	268	206	474
Total	639	377	1016

Sursa: Anuarul statistic al Republicii Socialiste România, 1977, D.C.S.,Bucureşti,p.464 - 65.

Ipoteza nulă poate fi formulată în mai multe feluri. Astfel, putem afirma că în alegerea tipului de liceu nu există nici o diferenţă între elevii din cele două sexe, sau că între

frecvenţele empirice şi cele teoretice nu există o diferenţă semnificativă. Dacă n-ar fi existat o preferinţă pe sexe pentru un tip sau altul de liceu, atunci ar fi trebuit să meargă la fiecare tip de liceu aceeaşi proporţie de elevi din fiecare sex. Ori, în exemplul nostru, frecvenţele teoretice diferă de cele empirice şi de aceea trebuie să verificăm dacă aceste diferenţe sînt semnificative sau nu. Pentru aceasta avem nevoie de un criteriu obiectiv şi acesta este criteriul χ^2.

Pentru aplicarea acestui criteriu este necesar să se determine, în prealabil, frecvenţele teoretice n_i', pe baza formulei (18.4), astfel:

$$\alpha = \frac{n_1 m_1}{n} = \frac{639.542}{1016} = 340,88385$$

$$\beta = \frac{n_2 m_1}{n} = \frac{377.542}{1016} = 201,11614$$

$$\gamma = \frac{n_1 m_2}{n} = \frac{639.474}{1016} = 298,11614$$

$$\partial = \frac{n_2 m_2}{n} = \frac{377.474}{1016} = 175,88385$$

Frecvenţele teoretice se mai pot obţine şi pe altă cale ceva mai simplă. Pornind de la expresia lui χ^2 şi efectuînd calculele cerute de semne, rezultă:

$$\chi^2 = \sum \frac{(n_i - n_i')^2}{n_i'} = \sum \frac{n_i^2 - 2\, n_i n_i' + n_i'^2}{n_i'} =$$

$$= \sum \frac{n_i^2}{n_i'} - 2 \sum n_i + \sum n_i' \qquad (18.11)$$

Ţinînd seama de relaţia $\sum n_i = \sum n_i'$, rezultă:

$$\chi^2 = \sum \frac{n_i^2}{n_i'} - n \qquad (18.12)$$

Pentru verificarea existenţei legăturii dintre sexe şi după categoria liceului, vom aplica criteriul χ^2. In acest scop calculele se sistematizează sub forma tabelului 18.5.

Tabelul 18.5. Calculul lui χ^2

	Frecvenţele empirice (n_i)	Frecvenţele teoretice (n_i')	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n-n_i')^2}{n_i'}$
a	371	341	30	900	2,6393
b	171	201	-30	900	4,4776
c	268	298	-30	900	3,0201
d	206	176	30	900	5,1136
Total	1016	1016	-	-	15,2506

Pentru determinarea lui χ^2 după formula (18.12), calculul se poate sistematiza cu ajutorul tabelului 18.6.

Tabelul 18.6. Calculul lui χ^2

	n_i^2	$\frac{n_i^2}{n_i'}$
a	137641	403,6393
b	29241	145,4776
c	71824	241,0201
d	42436	241,1136
Total	-	1031,2506

Deci $\chi^2 = 1031,2506 - 1016 = 15,2506$.

Calculul lui χ^2, în cazul unui tabel de contingenţă 2x2 se poate face şi pe baza formulei (18.10):

$$\chi^2 = \frac{n(a\,d-b\,c)^2}{n_1 n_2 m_1 m_2} = \frac{1016(371.206 - 171.268)^2}{371.206.171.268} = 15,2506$$

In cazul general al unui tabel de contingenţă k.h calculul lui χ^2 se face după formula:

$$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{h} \frac{(n_{ij}-n_{ij}')^2}{n_{ij}'} \qquad (18.13)$$

în care: $n'_{ij} = \frac{n_{i.} n_{.j}}{n}$

Efectuînd calculele în formula (18.13), rezultă:

$$\chi^2 = n \left(\sum_{i=1}^{k} \sum_{j=1}^{h} \frac{n_{ij}^2}{n_{i.} n_{.j}} - 1 \right) \qquad (18.14)$$

Din formula (18.14) se observă că pentru determinarea lui χ^2 nu mai este necesar să se calculeze în prealabil frecvenţele teoretice n'_{ij}.

O problemă importantă, ce trebuie rezolvată pentru aplicarea testului χ^2, este determinarea numărului <u>gradelor de libertate</u>.

La utilizarea testului χ^2, în cazul informaţiilor sistematizate cu ajutorul unui tabel de contingenţă, numărul gradelor de libertate depinde numai de numărul căsuţelor din tabel şi nu de numărul unităţilor statistice luate în calcul ca în cazul general. După cum am văzut, în cazul tabelelor de contingenţă 2 x 2, nu este necesar să se determine frecvenţele teoretice pentru toate căsuţele, ci este suficient să determinăm o singură frecvenţă teoretică şi toate celelalte vor fi determinate. De aceea, în cazul tabelelor de contingenţă 2 x 2 vom avea un singur grad de libertate.

In cazul general al unui tabel de contingenţă k x h numărul gradelor de libertate f va fi dat de formula:

$$f = (k - 1)(h - 1) \qquad (18.15)$$

Este clar că în cazul unui tabel de contingenţă 2 x 2 vom avea un număr al gradelor de libertate egal cu:

$$f = (2-1)(2-1) = 1$$

iar în cazul unui tabel de contingenţă 3 x 3 :

$$f = (3 - 1)(3 - 1) = 4$$

Pe baza numărului gradelor de libertate şi la un anumit nivel de semnificaţie se determină, din datele speciale, valoarea teoretică a lui χ^2. Astfel, în exemplul anterior (tabelul 18.5), la un nivel de semnificaţie $q_I = 0,001$ se găseşte o valoare $\chi^2_{1;0,999} = 10,8$. Deci valoarea lui χ^2 obţinută prin calcule $\chi^2_{calc.} = 15,25$ este mai mare decît valoarea tabelară, ceea ce înseamnă că ipoteza nulă se respinge cu un risc de a greşi într-un caz din 1000. Aceasta înseamnă că există o legătură reală între sex şi tipurile de licee luate în calcul şi anume băieţii preferă în mai mare măsură liceele de specialitate decît fetele.

Pentru verificarea existenţei asociaţiei, în cazul general, să considerăm distribuţia căsătoriilor după starea civilă a soţilor înainte de căsătorie în anul 1976 (tabelul 18.7).

Tabelul 18.7. Căsătorii după starea civilă a soţilor înainte de căsătorie (mii persoane)

Starea civilă a soţiei / Starea civilă a soţului	Necăsătorită	Văduvă	Divorţată şi nedeclarat	TOTAL
Necăsătorit	160,3	1,0	7,0	168,9
Văduv	1,7	2,4	1,8	5,9
Divorţat şi nedeclarat	11,4	1,5	8,1	21,0
T o t a l	173,4	4,9	17,5	195,8

Sursa: Anuarul statistic al R.S.R., D.C.S., p.78.

Pentru verificarea existenţei unei legături între starea civilă a soţilor înainte de căsătorie, vom proceda la verificarea ipotezei nule cu ajutorul testului χ^2. Ipoteza nulă presupune că între starea civilă a soţilor înainte de căsătorie nu ar exista nici un fel de legătură. Considerăm apoi un nivel de semnificaţie egal cu 0,001, iar numărul gradelor de libertate va fi egal cu: $f = (3-1)(3-1) = 4$.

Ca şi în cazul anterior calculul lui χ^2 în baza formulei (18.14) se va sistematiza cu ajutorul tabelului 18.8.

Tabelul 18.8. Calculul lui χ^2 în cazul unui tabel de contingenţă 3 x 3

Căsuţa	n_{ij}	n'_{ij}	n^2_{ij}	$\frac{n^2_{ij}}{n'_{ij}}$
n_{11}	160,3	149,6	25696,09	171,77
n_{12}	1,0	4,2	1,00	0,24
n_{13}	7,6	15,1	57,76	3,83
n_{12}	1,7	5,2	2,89	0,55
n_{22}	2,4	0,1	5,76	57,60
n_{23}	1,8	0,6	3,24	5,40
n_{31}	11,4	18,6	129,96	6,99
n_{32}	2,5	0,5	2,25	4,50
n_{33}	8,1	1,9	65,61	34,53
Total	195,8	195,8	-	285,41

$$\chi^2_{calc.} = \sum \frac{n^2_{ij}}{n'_{ij}} - n = 285,41 - 195,8 = 89,61$$

$$\chi^2_{tab} = \chi^2_{4;99,9} = 18,5$$

Deci $\chi^2_{calc.} > \chi^2_{tab.}$ înseamnă că ipoteza nulă se respinge, deci există o legătură între starea civilă a soţilor înainte de căsătorie, în sensul că se căsătoresc de regulă persoane cu aceeaşi stare civilă avută înainte de căsătorie.

18.3. Măsurarea intensităţii asociaţiei

In continuare vom prezenta unii indicatori ai intensităţii asociaţiei cu ajutorul cărora să se poată caracteriza numeric intensitatea legăturii dintre variabilele exprimate prin cuvinte ca şi în cazul caracteristicilor cantitative.

In cazul unui tabel de contingenţă 2 x 2 (tabelul 18.2) se pot ivi două situaţii opuse:

- <u>independenţă totală</u> cînd are loc relaţia:

$$ad = bc \qquad (18.16)$$

- <u>asociaţie pozitivă completă</u> cînd $bc = 0$. Aceasta se poate întîmpla în una dintre următoarele situaţii:

$$\begin{array}{lllll} 1) & b = 0 & şi & c \neq 0 & \\ 2) & c = 0 & şi & b \neq 0 & \qquad (18.17) \\ 3) & b = 0 & şi & c = 0 & \end{array}$$

- <u>asociaţia negativă completă</u> cînd $ad = 0$, adică:

$$\begin{array}{lllll} 1) & c = 0 & şi & d \neq 0 & \\ 2) & d = 0 & şi & a \neq 0 & \qquad (18.18) \\ 3) & a = 0 & şi & d = 0 & \end{array}$$

Dintre aceste situaţii notăm faptul că în cel de al treilea caz asociaţia este mai completă decît în primele două, atît în cazul asociaţiei pozitive, cît şi în cazul celei negative.

In literatura de specialitate se cunosc o mulţime de indicatori ai intensităţii asociaţiei dintre variabile, dar aici vor fi prezentaţi numai cei mai reprezentativi.

<u>Coeficientul de asociaţie al lui G.U. Yule</u> este definit de relaţia:

$$Q = \frac{ad - bc}{ad + bc} \qquad (18.19)$$

Din formula (18.19) rezultă: $-1 \leq Q \leq 1$

Astfel: $Q = 0$ dacă $ad = bc$, deci în cazul cînd variabilele luate în studiu sînt independente;

$Q = 1$ dacă $bc = 0$ adică între varibile există o asociaţie pozitivă completă.

$Q = -1$ dacă $ad = 0$, cînd între variabile există o asociaţie negativă completă.

Acest indicator prezintă dezavantajul că nu face nici o

distincţie între cele trei cazuri posibile de asociaţie completă pozitivă sau negativă.

Aplicat la exemplul anterior din tabelul 18.4, rezultă:

$$Q = \frac{371.206 - 171.268}{371.206 + 171.268} = \frac{30598}{122254} = 0,25$$

$Q = 0,25$ indică că între sexe şi tipul liceului preferat există o asociaţie pozitivă slabă.

Un alt indicator de măsurare a intensităţii asociaţiei, în cazul unui tabel de asociaţie 2 x 2 , este coeficientul de coligaţie al lui Yule , care are aceleaşi proprietăţi şi deci aceleaşi deficienţe ca şi coeficientul său de asociaţie.

Coeficientul de coligaţie Y este dat de relaţia:

$$Y = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \qquad (18.20)$$

In exemplul anterior:

$$Y = \frac{\sqrt{371.206} - \sqrt{171.268}}{\sqrt{371.206} + \sqrt{171.268}} = \frac{62,38}{490,52} = 0,127$$

Pătratul de contingenţă Φ^2 , a fost propus de către statisticianul englez K. Pearson şi este definit de formula:

$$\Phi^2 = \frac{\chi^2}{n} \quad \frac{(ad - bc)^2}{m_1 n_1 m_2 n_2} \qquad (18.21)$$

sau

$$\Phi^2 = \frac{ad - bc}{\sqrt{m_1 m_2 n_1 n_2}} \qquad (18.22)$$

Din formula (18.22) rezultă că:

$$-1 \leq \Phi \leq 1 \qquad (18.23)$$

$\Phi = 0$ când $ad - bc$ şi corespunde independenţei dintre cele două caracteristici;

$\Phi = \pm 1$ când : $a = d = 0$

sau $b = c = 0$

Pătratul de contingenţă este egal cu ± 1 numai în cazul asociaţiilor pozitive şi respectiv negative complete.

Aplicînd formula (18.22) la exemplul din tabelul 18.4, rezultă:

$$\Phi = \frac{371.206 - 171.268}{\sqrt{542.474.639.377}} = \frac{30598}{248776,81} = 0,12299$$

Menţionăm faptul că pentru aceleaşi date, deci pentru aceeaşi intensitate a asociaţiei, cei trei indicatori Q, Y şi Φ dau rezultate diferite. Acest lucru se verifică şi în exemplul nostru anterior. De aceea, pentru compararea intensităţii asociaţiei dintre perechi diferite de variabile, este necesar să se utilizeze acelaşi indicator de intensitate.

In cazul general al tabelelor de contingenţă k x h, intensitatea asociaţiei dintre două caracteristici se determină cu ajutorul coeficientului de contingenţă C al lui K. Pearson, care se bazează pe diferenţele dintre frecvenţele reale şi cele teoretice $n_{ij} - n'_{ij}$ calculate pentru toate căsuţele i şi j din cadrul tabelului.

Coeficientul de contingenţă este dat de următoarea formulă:

$$C = \frac{\Phi^2}{1 + \Phi^2} = \frac{\chi^2}{n + \chi^2} \qquad (18.24)$$

unde:

$$\Phi^2 = \frac{1}{n} \sum\sum \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}} = \frac{\chi^2}{n} \qquad (18.25)$$

$C = 0$ cînd $\chi^2 = 0$, iar cele două caracteristici sînt independente. Valoarea lui C creşte pe măsură ce creşte χ^2. K. Pearson a demonstrat că C reprezintă o estimaţie a coeficientului de corelaţie R dintr-o colectivitate.

Aplicat la exemplul din tabelul 18.8, unde $\chi^2 = 89,61$, iar $n = 195,8$, rezultă:

$$C = \frac{89.61}{195,8+89,61} = 0,313969$$

$C = 0,3139$ indică faptul că între starea civilă a soţilor înainte de căsătorie există o legătură directă - se căsătoresc, în cea mai mare parte, persoane care au avut aceeaşi stare civilă înainte de căsătorie - şi de intensitate relativ slabă.

Un alt indicator al intensităţii asociaţiei îl constituie coeficientul lui Ciuprov - notat cu T şi definit de relaţia:

$$T^2 = \frac{\Phi^2}{\sqrt{(k-1)(h-1)}} \qquad (18.26)$$

T^2 faţă de C prezintă avantajul că poate atinge valoarea limită superioară, atunci cînd χ^2 creşte.

Calculînd acest indicator, pe baza datelor din exemplul anterior, se obţine:

$$\Phi^2 = \frac{\chi^2}{h} = \frac{89,61}{195,61} = 0,4581054$$

$$T^2 = \frac{0,4581054}{2} = 0,2290$$

$T^2 \simeq 0,23$ evidenţiază faptul că între variabile există o legătură directă şi de intensitate redusă.

19. INDICATORII SERIILOR DINAMICE

Timpul şi spaţiul reprezintă cele două forme fundamentale de existenţă a materiei. Dezvoltarea oricărui fenomen din natură sau din societate are loc într-un anumit spaţiu care îşi imprimă asupra fenomenelor anumite trăsături specifice şi într-un anumit moment sau interval de timp, caracteristică care, de asemenea, imprimă fenomenelor anumite trăsături. Corespunzător acestor două caracteristici de bază, în care are loc naşterea şi dezvoltarea fenomenelor, se disting în statistică două categorii de distribuţii şi anume: distribuţii (serii) de spaţiu şi distribuţii (serii) de timp.

Termenii seriilor dinamice prezintă nivelul de dezvoltare a fenomenelor în anumite intervale de timp bine determinate şi, de regulă, egale între ele pentru a se putea face comparaţie între intervale sau nivelul atins în diferite momente de timp din dezvoltarea lor. Intervalele dintre momente pot fi egale sau neegale între ele.

Corespunzător categoriilor principale de mărimi folosite în statistică, indicatorii care se utilizează în caracterizarea seriilor dinamice se pot împărţi în următoarele categorii:

- indicatori absoluţi: - nivelul absolut;
  - sporul absolut;
- indicatori relativi: - ritmul variaţiei totale;
  - ritmul sporului;
- indicatori medii : - sporul mediu absolut
  - ritmul mediu al variaţiei
  - ritmul mediu al sporului.

In dezvoltarea fenomenelor în timp sub influenţa diferiţilor factori care acţionează asupra lor se pot evidenţia mai multe componente:

- o direcţie fundamentală de evoluţie dată de influenţa factorilor esenţiali ai dezvoltării, numită şi tendinţa de dezvoltare sau, mai simplu, trend;

- variaţii sau oscilaţii ciclice sezoniere determinate de unii factori naturali cu acţiune sezonieră sau factori sociali;

- variaţii întîmplătoare determinate de influenţa unor factori întîmplători.

Analiza variaţiei fenomenelor în timp înseamnă separarea celor trei componente ale seriilor dinamice.

19.1. Indicatorii nivelului şi sporului absolut al seriilor dinamice

Pentru exemplificarea metodelor de calcul şi analiză statistică a indicatorilor nivelului şi sporului absolut al seriilor dinamice (de momente sau intervale), considerăm seria dinamică din tabelul 19.1.

Tabelul 19.1. Indicatorii seriilor dinamice exprimaţi în mărimi absolute

Momentul sau intervalul (t_i)	Nivelurile atinse de un anumit fenomen (y_i)	Sporul absolut	
		Cu bază fixă ($\Delta_{y_{i/0}}$)	Cu bază mobilă ($\Delta_{y_{i/i-i}}$)
1	2	3	4
t_0	y_0	-	-
t_1	y_1	$\Delta_{y_{1/0}}$	$\Delta_{y_{1/0}}$
t_2	y_2	$\Delta_{y_{2/0}}$	$\Delta_{y_{2/1}}$
⋮	⋮	⋮	⋮
t_i	y_i	$\Delta_{y_{i/0}}$	$\Delta_{y_{i/i-1}}$
⋮	⋮	⋮	⋮
t_n	y_n	$\Delta_{y_{n/0}}$	$\Delta_{y_{n/n-1}}$
Total	$\sum_{i=0}^{n} y_i$	-	$\Delta_{y_{n/0}}$

Sporul sau excedentul absolut măsoară în mărimi absolute variaţia unui fenomen dintr-un anumit interval de timp şi se calculează cu o diferenţă între nivelul atins de un fenomen la un moment dat şi nivelul aceluiaşi fenomen dintr-un alt moment anterior, considerat ca bază de comparaţie. Termenul ales ca bază de comparaţie trebuie să reprezinte un nivel normal din dezvoltarea fenomenului luat în calcul, supus într-o măsură cît mai mică influenţei factorilor perturabatori ai dezvoltării. In funcţie de modul de alegere a bazei de comparaţie într-o serie dinamică se disting două categorii de spor absolut:

- un spor absolut cu bază fixă - notat cu $\Delta_{y_{i/o}}$
- un spor absolut cu bază mobilă - notat cu $\Delta_{y_{i/i-1}}$

Sporul absolut cu bază fixă se caracterizează prin faptul că nivelul considerat ca bază rămîne acelaşi pentru întreaga serie. Dacă considerăm ca nivel de bază primul termen al unei serii dinamice y_o, atunci sporul absolut cu bază fixă se poate calcula cu ajutorul formulei:

$$\Delta_{y_{i/o}} = y_i - y_o \qquad i = 0,1,2,\ldots,n \qquad (19.1)$$

In domeniul economic se consideră ca bază de comparaţie un nivel din dezvoltarea unui fenomen cu o semnificaţie aparte. De exemplu, pentru a caracteriza creşterea producţiei unor produse industriale principale pe durata unui plan cincinal, se compară nivelurile atinse în diferite etape ale cincinalului în curs cu nivelul aceluiaşi fenomen atins la sfîrşitul cincinalului precedent, sau pentru a scoate în evidenţă succesele obţinute de poporul nostru în anii construcţiei socialiste, se face comparaţie cu anul 1938, anul cu nivelul cel mai ridicat de dezvoltare economică a României înainte de eliberare. In toate cazurile, alegerea bazei de comparaţie nu trebuie să se facă la întîmplare, ci numai în urma unei analize temeinice a evoluţiei fenomenelor în timp şi în funcţie de scopul (cunoaştere ştiinţifică) urmărit în fiecare caz în parte. Precizăm, de asemenea, că baza de comparaţie nu trebuie să fie prea mult depărtată în timp de nivelul care se compară, deoarece dezvoltarea în timp înseamnă nu numai o creştere

(sau scădere) cantitativă a aceleiaşi calităţi, ci şi modificări calitative, de esenţă, uneori chiar foarte importante, care modifică însăşi natura fenomenelor şi le fac să nu mai fie întrutotul comparabile.

Sporul absolut cu bază mobilă se caracterizează prin faptul că baza de comparaţie se schimbă în mod continuu, de la un termen la altul, considerîndu-se de fiecare dată drept bază nivelul precedent. Astfel, se caracterizează variaţia fenomenelor de la o zi la alta, de la o săptămînă la alta, de la o lună la alta, de la un an la altul, de la un cincinal la altul etc.

Formula de calcul al sporului absolut cu bază mobilă sau în lanţ, cum i se mai spune, este:

$$\Delta_{y_{i/i-1}} = y_i - y_{i-1} \qquad i = 0,1,2,\ldots,n \qquad (19.2)$$

Pentru ilustrarea calculului şi analizei statistice a sporului absolut al fenomenelor, să considerăm seria dinamică din tabelul 19.2, reprezentînd creşterea producţiei de energie electrică în România în anii 1938, 1950, 1960 şi 1965-1976.

Între sporurile cu bază fixă şi sporurile cu bază mobilă există anumite relaţii bine determinate, care permit trecerea de la o categorie de sporuri la cealaltă fără a se recurge la seria dinamică pe baza căreia au fost calculate aceste sporuri.

Tabelul 19.2. Producţia de energie electrică în anii 1938, 1950, 1960 şi 1965 - 1976

Anul	Milioane kwh (y_i)	Sporul absolut					
		Cu baza fixă :					Cu bază mobilă
		y_{1938}	y_{1950}	y_{1960}	y_{1965}	y_{1970}	
1	2	3	4	5	6	7	8
1938	1130	0	-983	-6520	-16085	-	-
1950	2113	983	0	-5537	-15102	-	983
1960	7650	6520	5537	0	- 9565	-	5537
1965	17215	16085	15102	9565	0	-	9565
1966	20806	19676	18693	13156	3591	-	3591
1967	24769	23639	22656	17119	7554	-	3963
1968	27828	26698	25715	20178	10613	-	3059
1969	31509	30379	29396	23859	14294	-	3681
1970	35088	33958	32975	27438	17873	0	3579

Tabelul 19.2 (continuare)

1	2	3	4	5	6	7	8
1971	39454	38324	37341	31804	22239	4366	4366
1972	43439	42309	41326	35789	26224	8351	3985
1973	46779	45649	44666	39129	29564	11691	3340
1974	49063	47933	46950	41413	31848	13975	2284
1975	53721	52591	51608	46071	36506	18633	4658
1976	58266	57136	56153	50616	41051	23178	4545

Sursa: Anuarul statistic al R.S.R., 1971, DCS, Bucureşti, p.198-199
Anuarul statistic al R.S.R., 1976, DCS, Bucureşti, p. 112-113.
Anuarul statistic al R.S.R., 1977, DCS, Bucureşti, p. 152-153.

1. Suma sporurilor cu bază mobilă, dintr-o anumită perioadă de timp, este egală cu sporul de bază fixă al întregii perioade, adică:

$$\sum_{i=1}^{n} \Delta_{y_{i/i-1}} = \Delta_{y_{n/o}} \qquad (19.3)$$

Relaţia (19.3) se demonstrează astfel:

$$\Delta_{y_{1/o}} = y_1 - y_o$$

$$\Delta_{y_{2/1}} = y_2 - y_1$$

$$\Delta_{y_{3/2}} = y_3 - y_2$$

$$\cdots\cdots\cdots\cdots$$

$$\Delta_{y_{i/i-1}} = y_i - y_{i-1}$$

$$\cdots\cdots\cdots\cdots$$

$$\Delta_{y_{n/n-1}} = y_n - y_{n-1}$$

Insumînd aceste egalităţi rezultă:

$$\sum_{i=1}^{n} \Delta_{y_{i/i-1}} = (y_1-y_o)+(y_2-y_1)+(y_3-y_2)+\ldots+(y_i-y_{i-1})+$$

$$+ \ldots +(y_n-y_{n-1}) = y_n-y_o = \Delta_{y_{n/o}}$$

In exemplul anterior totalul coloanei 8 din tabelul 19.2 este egal cu sporul cu bază fixă (y_{1938} = bază) al perioadei 1938-1976 (ultimul rînd de pe coloana 3).

2. Diferenţa dintre două sporuri absolute cu bază fixă consecutive este egală cu sporul cu bază mobilă corespunzător:

$$\Delta_{y_{i/o}} - \Delta_{y_{i-1/o}} = \Delta_{y_{i/i-1}} \qquad i = 1,2,\ldots,n \qquad (19.4)$$

Relaţia (19.4) se verifică imediat:

$$\Delta_{y_{1/o}} - \Delta_{y_{o/o}} = (y_1 - y_o) - (y_o - y_o) = y_1 - y_o = \Delta_{y_{1/o}}$$

$$\Delta_{y_{2/o}} - \Delta_{y_{1/o}} = (y_2 - y_o) - (y_1 - y_o) = y_2 - y_1 = \Delta_{y_{2/1}}$$

.....

$$\Delta_{y_{i/o}} - \Delta_{y_{i-1/o}} = (y_i - y_o) - (y_{i-1} - y_o) = y_i - y_{i-1} = \Delta_{y_{i/i-1}}$$

.....

$$\Delta_{y_{n/o}} - \Delta_{y_{n-1/o}} = (y_n - y_o) - (y_{n-1} - y_o) = y_n - y_{n-1} = \Delta_{y_{n/n-1}}$$

Un alt indicator al variaţiei fenomenelor, exprimat în aceleaşi unităţi de măsură ca şi fenomenul însuşi, este sporul mediu ($\bar{\Delta}_y$). Sporul mediu arată în mărimi absolute cît de mare ar fi fost variaţia (creşterea sau descreşterea) unui fenomen în perioade mai mici de timp dar egale şi cuprinse în cadrul perioadei mai mari luate în calcul.

Sporul mediu se poate calcula ca o medie aritmetică simplă a sporurilor cu bază mobilă, după formula:

$$\bar{\Delta}_y = \frac{\sum_{i=1}^{n} \Delta_{y_{i/i-1}}}{n} \qquad (19.5)$$

Sporul mediu anual de creştere a producţiei de energie energie electrică pentru perioada 1938-1976 ca şi pentru diferite subperioade, a avut valorile următoare:

$$\bar{\Delta}_{y_{1938-1976}} = \frac{58266 - 1130}{38} = 1503,5 \text{ milioane kwh}$$

Aceasta înseamnă că în perioada 1938-1976 producţia de energie electrică a crescut cu 1503,5 milioane kwh în fiecare an.

Pentru diferite subperioade s-au obţinut sporurile medii anuale, astfel:

$$\bar{\Delta}_{y_{1950-1960}} = \frac{5537}{10} = 553,7 \text{ milioane kwh}$$

$$\bar{\Delta}_{y_{1960-1970}} = \frac{27438}{10} = 2743,8 \text{ milioane kwh}$$

$$\bar{\Delta}_{y_{1965-1970}} = \frac{17873}{5} = 3574,6 \text{ milioane kwh}$$

$$\bar{\Delta}_{y_{1970-1975}} = \frac{18633}{5} = 3726,6 \text{ milioane kwh}$$

In baza relaţiei (19.3) formula sporului mediu devine:

$$\bar{\Delta}_y = \frac{\overset{\Delta}{y_{n/0}}}{n} \qquad (19.6)$$

Relaţia (19.6) arată că sporul mediu se poate calcula şi ca un raport între sporul cu bază fixă al întregii perioade şi numărul termenilor seriei, minus 1. În formula (19.6), ca şi în cele anterioare, s-a considerat că seria pentru care se calculează indicatorii sporului este formată din n+1 termeni.

Dacă în relaţia (19.6) se substituie sporul cu bază fixă al întregii perioade cu expresia lui din (19.1), rezultă formula uzuală de calcul al sporului mediu:

$$\bar{\Delta}_y = \frac{y_n - y_0}{n} \qquad (19.7)$$

Din formula (19.7) se observă că sporul mediu depinde numai de mărimea termenilor extremi ai unei serii şi de numărul lor, făcînd abstracţie de toţi termneii intermediari ai seriei. Aceasta înseamnă că metoda clasică de determinare a sporului mediu exagerează artificial importanţa primului şi a ultimului termen ai unei serii şi neglijează pe toţi ceilalţi termeni.

Constatarea aceasta reprezintă principala deficienţă a metodei obişnuite de calcul al sporului mediu - deficienţă care-i limitează sfera de aplicabilitate practică numai la seriile dinamice în care termenii extremi reprezintă niveluri normale în dezvoltarea fenomenului studiat. In caz contrar, toate rezultatele, obţinute prin prelucrarea statistică ulterioară, se vor abate faţă de realitate aproximativ în aceeaşi măsură în care termenii y_0 şi y_n , sau ambii termeni se abat de la dezvoltarea normală a fenomenului.

De aceea, ori de cîte ori se constată că unul sau ambii termeni extremi se abat de la dezvoltarea normală a fenomenului cercetat, pentru a se înlătura deficienţa arătată, considerăm că este mai potrivit să se calculeze sporul mediu cu ajutorul metodei celor mai mici pătrate, care ţine seama de toate nivelurile individuale ale seriei.

Să presupunem că dezvoltarea fenomenului cercetat are loc în mod liniar, după funcţia:

$$y = a + bt$$

Sporul mediu poate fi estimat ca un raport între creşterea funcţiei şi creşterea argumentului, astfel:

$$y + \Delta_y = a + b(t + \Delta_t)$$

sau

$$y + \Delta_y = a + bt + b\Delta_t$$

Scăzînd din valoarea funcţiei în momentul $t + \Delta t$ valoarea ei în momentul t , rezultă:

$$\Delta_y = b\Delta_t$$

iar

$$b = \frac{\Delta_y}{\Delta_t} \qquad (19.8)$$

Din relaţia (19.8) se observă că parametrul b din ecuaţia unei drepte reprezintă viteza de creştere a funcţiei liniare sau sporul mediu.

Sporul mediu este un indicator sintetic al variaţiei fenomenelor cu importante aplicaţii în analiza şi prognoza feno-

menelor pe baza seriilor dinamice.

19.2. Ritmul variaţiei totale a fenomenelor

Un alt indicator important al variaţiei fenomenelor în timp îl reprezintă viteza de variaţie, exprimată în mărimi relative, numit ritm de variaţie. Ca orice mărime relativă şi ritmul variaţiei în timp a fenomenelor se obţine prin compararea nivelului unui fenomen, atins într-un anumit moment, cu un alt nivel anterior al aceluiaşi fenomen, considerat ca bază de comparaţie. In funcţie de modul de alegere a bazei de comparaţie, ritmul variaţiei fenomenelor este de două feluri:

- ritmul variaţiei fenomenelor cu bază fixă ($R_{y_i/o}$);
- ritmul variaţiei fenomenelor cu bază mobilă ($R_{y_i/i-1}$).

Ritmul variaţiei cu bază fixă arată de cîte ori a crescut (sau scăzut) nivelul unui anumit fenomen în decursul unei anumite perioade de timp şi se calculează după formula:

$$R_{y_i/o} = \frac{y_i}{y_o} \qquad i = 0,1,2,\ldots,n \qquad (19.9)$$

Ritmul variaţiei cu bază mobilă se calculează după formula:

$$R_{y_i/i-1} = \frac{y_i}{y_{i-1}} \qquad i = 0,1,2,\ldots,n \qquad (19.10)$$

şi arată de cîte ori a crescut sau a scăzut nivelul unui anumit fenomen într-un anumit moment sau interval de timp faţă de un moment sau perioadă anterioară.

Consideraţiile privitoare la alegerea bazei de comparaţie expuse cu privire la sporul absolut sînt întrutotul aplicabile şi în cazul ritmului variaţiei.

Relaţiile dintre ritmurile variaţiei cu bază fică şi cele cu bază mobilă se pot enunţa astfel:

1. Produsul ritmurilor variaţiei cu bază mobilă este egal cu ritmul variaţiei cu bază fixă al întregii perioade:

$$\prod_1^n R_{y\ i/i-1} = R_{y\ n/o} \qquad (19.11)$$

Intr-adevăr

$$\prod_1^n R_{y\ i/i-1} = R_{y\ 1/0} \cdot R_{y\ 2/1} \cdot R_{y\ 3/2} \cdots R_{y\ i/i-1} \cdots R_{y\ n/n-1} =$$

$$= \frac{y_1}{y_0} \cdot \frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \cdots \frac{y_i}{y_{i-1}} \cdots \frac{y_n}{y_{n-1}} = \frac{y_n}{y_0} = R_{y\ n/0}$$

2. Raportul între două ritmuri ale variaţiei cu bază fixă este egal cu ritmul cu bază mobilă corespunzător:

$$R_{y\ i/0} : R_{y\ i-1/0} = R_{y\ i/i-1} \qquad (19.12)$$

Astfel

$$R_{y\ 1/0} : R_{y\ \%} = \frac{y_1}{y_0} : \frac{y_0}{y_0} = \frac{y_1}{y_0} = R_{y\ 1/0}$$

$$R_{y\ 2/0} : R_{y\ 1/0} = \frac{y_2}{y_0} : \frac{y_1}{y_0} = \frac{y_2}{y_1} = R_{y\ 2/1}$$

.....

$$R_{y\ i/0} : R_{y\ i-1/0} = \frac{y_i}{y_0} : \frac{y_{i-1}}{y_0} = \frac{y_i}{y_{i-1}} = R_{y\ i/i-1}$$

.....

$$R_{y\ n/0} : R_{y\ n-1/0} = \frac{y_n}{y_0} : \frac{y_{n-1}}{y_0} = \frac{y_n}{y_{n-1}} = R_{y\ n/n-1}$$

Pentru exemplificare, considerăm seria din tabelul 19.2.

Tabelul 19.3. Producţia de energie electrică în anii 1938, 1950, 1960 şi 1965-1976

Anul	Mili-oane kwh	Ritmul variaţiei cu bază fixă :					Ritmul variaţiei cu bază mobilă
		y_{1938}	y_{1950}	y_{1960}	y_{1965}	y_{1970}	
1	2	3	4	5	6	7	8
1938	1130	1,0000	-	-	-	-	-
1950	2113	1,8699	1,0000	-	-	-	1,8699
1960	7650	6,7699	3,6204	1,0000	-	-	3,6204
1965	17215	15,2345	8,1472	2,2503	1,0000	-	2,2503
1966	20806	18,4124	9,8467	2,7197	1,2086	-	1,2086
1967	24769	21,9195	11,7222	3,2378	1,4388	-	1,1905
1968	27828	24,6265	13,1699	3,6376	1,6165	-	1,1235
1969	31509	27,8841	14,9120	4,1188	1,8303	-	1,1323
1970	35088	31,0513	16,6058	4,5867	2,0382	1,0000	1,1136
1971	39454	34,9150	18,6720	5,1574	2,2918	1,1244	1,1244
1972	43439	38,4416	20,5580	5,6783	2,5233	1,2380	1,1010
1973	46779	41,3973	22,1387	6,1149	2,7173	1,3332	1,0769
1974	49063	43,4186	23,2196	6,4135	2,8506	1,3983	1,0488
1975	53721	47,5407	25,4240	7,0224	3,1206	1,5310	1,0949
1976	58266	51,5628	27,5750	7,6165	3,3846	1,6606	1,0846

19.3. Ritmul creşterii fenomenelor

Ritmul creşterii fenomenelor exprimă mărimea creşterii (sau scăderii) în decursul unei anumite perioade de timp faţă de nivelul considerat de bază. După modul de alegere a bazei de comparaţie se disting două feluri de ritmuri ale creşterii, numite uneori şi ritmuri ale sporului:

- ritmul creşterii cu bază fixă ($r_{y\ i/o}$)
- ritmul creşterii cu bază mobilă ($r_{y\ i/i-1}$)

Ritmul creşterii cu bază fixă se calculează prin raportarea sporului cu bază fixă la nivelul fenomenului considerat ca bază de comparaţie, după formula:

$$r_{y\ i/o} = \frac{\Delta_{y\ i/o}}{y_o} \qquad (19.13)$$

sau

$$r_{y\ i/o} = \frac{y_i - y_o}{y_o}$$

Ritmul creşterii cu bază mobilă se poate calcula prin raportarea sporului cu bază mobilă la nivelul din perioada de bază, astfel:

$$r_{y\ i/i-1} = \frac{\Delta_{y\ i/i-1}}{y_{i-1}} \qquad (19.14)$$

sau

$$r_{y\ i/i-1} = \frac{y_i - y_{i-1}}{y_{i-1}}$$

Intre ritmul variaţiei totale şi ritmul creşterii există o relaţie bine determinată şi anume: ritmul creşterii este egal cu ritmul variaţiei totale minus 1 sau 100, în funcţie de modul de exprimare a acestor ritmuri sub formă de coeficient sau în procente. Astfel:

$$r_{y\ i/o} = R_{y\ i/o} - 1 \qquad (19.15)$$

deoarece

$$r_{y\ i/o} = \frac{y_i - y_o}{y_o} = \frac{y_i}{y_o} - 1 = R_{y\ i/o} - 1$$

şi

$$r_{y\ i/i-1} = R_{y\ i/i-1} - 1$$

deoarece

$$r_{y\ i/i-1} = \frac{y_i - y_{i-1}}{y_{i-1}} = \frac{y_i}{y_{i-1}} - 1 = R_{y\ i/i-1} - 1$$

Sintetic, indicatorii relativi ai seriilor dinamice se pot sistematiza cu ajutorul tabelului 19.4.

Tabelul 19.4. Indicatorii relativi ai seriilor dinamice

Momentul sau intervalul de timp (t_i)	Nivelul de dezvoltare atins de un anumit fenomen (y_i)	Ritmul variației		Ritmul creșterii	
		Cu bază fixă $R_{y\ i/o}$	Cu bază mobilă $R_{y\ i/i-1}$	Cu bază fixă $r_{y\ i/o}$	Cu bază mobilă $r_{y\ i/i-1}$
t_o	y_o	$R_{y\ o/o}$	-	$r_{y\ o/o}$	-
t_1	y_1	$R_{y\ 1/o}$	$R_{y\ 1/o}$	$r_{y\ 1/o}$	$r_{y\ 1/o}$
t_2	y_2	$R_{y\ 2/o}$	$R_{y\ 2/1}$	$r_{y\ 2/o}$	$r_{y\ 2/1}$
⋮	⋮	⋮	⋮	⋮	⋮
t_i	y_i	$R_{y\ i/o}$	$R_{y\ i/i-1}$	$r_{y\ i/o}$	$r_{y\ i/i-1}$
⋮	⋮	⋮	⋮	⋮	⋮
t_n	y_n	$R_{y\ n/o}$	$R_{y\ n/n-1}$	$r_{y\ n/o}$	$r_{y\ n/n-1}$
	$\sum_{o}^{n} y_i$	-	$\prod_{1}^{n} R_{y\ i/i-1}$	-	

19.3.1. Ritmul mediu de creștere

Ritmul mediu de creștere exprimă viteza medie relativă de creștere a unui termen în decursul unei anumite perioade de timp. Astfel, se poate calcula ritmul mediu anual de creștere a unui fenomen în decursul unui cincinal sau a unei perioade de 10 ani etc. Ritmul mediu anual măsoară creșterea medie anuală a unui fenomen în ipoteza că în decursul întregii perioade fenomenul s-ar fi dezvoltat cu aceeași creștere relativă.

In teoria și practica statistică se cunosc mai multe metode de determinare a ritmului mediu de creștere a fenomenelor, dintre care, mai des întrebuințate sînt: metoda mediei geometrice, metoda mediei parabolice și metoda trendului exponențial.

19.3.1.1. Metoda mediei geometrice

In ipoteza dezvoltării unui fenomen cu ritmuri de variaţie cu bază mobilă constante, după modelul unei progresii geometrice, se poate determina ritmul mediu de creştere ca o raţie a progresiei geometrice. In această ipoteză termenii seriei dinamice se pot scrie astfel:

$$y_0 = y_0 \bar{R}_y^0$$

$$y_1 = y_0 \bar{R}_y$$

$$y_2 = y_0 \bar{R}_y^2$$

$$\cdots\cdots\cdots$$

$$y_i = y_0 \bar{R}_y^i$$

$$\cdots\cdots\cdots$$

$$y_n = y_0 \bar{R}_y^n \qquad (19.17)$$

Ritmul mediu de creştere ($\bar{R}_y$) se poate determina pe baza relaţiei (19.17) astfel:

$$\bar{R}_y^n = \frac{y_n}{y_0}$$

iar

$$\bar{R}_y = \sqrt[n]{\frac{y_n}{y_0}} = \sqrt[n]{R_{n/0}} \qquad (19.18)$$

Relaţia (19.18) arată că ritmul mediu de creşte a unui anumit fenomen se poate determina prin extragerea rădăcinii de un ordin egal cu numărul termenilor seriei minus 1 din ritmul de variaţie cu bază fixă al întregii perioade.

Ţinînd seama de relaţia (19.11) formula ritmului mediu de creştere devine:

$$\bar{R}_y = \sqrt[n]{\prod_1^n R_{i/i-1}} \qquad (19.19)$$

Din relaţia (19.19) rezultă că ritmul mediu de creştere al fenomenelor se poate determina ca o medie geometrică a ritmurilor de variaţie cu bază mobilă.

Prin logaritmare relaţia (19.18) devine:

$$\log \bar{R}_y = \frac{1}{n} (\log y_n - \log y_0)$$

şi

$$\bar{R}_y = \text{antilog} \left[\frac{1}{n} (\log y_n - \log y_0) \right]$$

In exemplul anterior (tabelul 19.3) vom avea:

- pentru perioada 1938-1976:

$$\bar{R}_y = \sqrt[38]{51,5628} = 1,109$$

- pentru perioada 1950-1976:

$$\bar{R}_y = \sqrt[26]{27,575} = 1,136$$

- pentru perioada 1960-1976:

$$\bar{R}_y = \sqrt[16]{7,6165} = 1,135$$

- pentru perioada 1965-1976:

$$\bar{R}_y = \sqrt[11]{3,3846} = 1,117$$

- pentru perioada 1970-1976 :

$$\bar{R}_y = \sqrt[6]{1,6606} = 1,089$$

- pentru perioada 1970-1975:

$$\bar{R}_y = \sqrt[5]{1,5310} = 1,089$$

Rezultă că producţia de energie electrică a României a crescut, în medie în fiecare an al perioadei 1938-1976, de 1,113 ori. Datorită faptului că baza luată în calcul, anul 1938, a fost foarte mică, comparativ cu nivelurile mai recente, ritmul mediu al variaţiei a fost mai ridicat pe ansamblul perioadei şi a înregistrat o scădere lentă în anii ce au urmat. Este clar însă că scăderea ritmului de variaţie pentru perioadele mai recente se datorează în special luării în calcul a unei baze tot mai mari.

Ritmul mediu se poate calcula pentru perioade mai mari sau mai mici de timp. Dacă se cunosc ritmurile medii de creştere ($\bar{R}_i$) ale unui fenomen în diferite subperioade succesive de timp, atunci pe baza acestora se poate determina ritmul mediu de creştere pentru întreaga perioadă ($\bar{R}$), ca o medie geometrică ponderată a ritmurilor de creştere, pe subperioade, astfel:

$$\bar{R} = \sqrt[\sum n_i]{\bar{R}_1^{n_1} \cdot \bar{R}_2^{n_2} \dots \bar{R}_i^{n_i} \dots \bar{R}_k^{n_k}} \qquad (19.30)$$

In cadrul exemplului din tabelul 13.3, ritmul mediu anual de variaţie al producţiei de energie electrică pentru întreaga perioadă 1950-1975 se poate calcula cu ajutorul ritmurilor medii obţinute pentru diferite perioade egale (cincinale). Astfel:

- pentru perioada 1950 - 1975:

$$\bar{R} = \sqrt[25]{25,434} = 1,138$$

- pentru perioada 1950 - 1955:

$$\bar{R} = \sqrt[5]{\frac{4340}{2113}} = 1,155$$

- pentru perioada 1955 - 1960:

$$\bar{R} = \sqrt[5]{\frac{7650}{4340}} = 1,12$$

- pentru perioada 1960 - 1965:

$$\bar{R} = \sqrt[5]{2,2503} = 1,176$$

- pentru perioada 1965 - 1970:

$$\bar{R} = \sqrt[5]{2,0382} = 1,153$$

- pentru perioada 1970 - 1975:

$$\bar{R} = \sqrt[5]{1,5310} = 1,089$$

Ritmul mediu anual de variaţie pentru întreaga perioadă se poate determina ca o medie geometrică simplă a ritmurilor de variaţie pe cincinale, deoarece acestea sînt egale:

Scanned with OKEN Scanner

$$\bar{R} = \sqrt[5]{1,155 \cdot 1,120 \cdot 1,176 \cdot 1,153 \cdot 1,089} = \sqrt[5]{1,91014} = 1,138$$

In cazul calculării ritmului mediu al variației pentru întreaga perioadă de timp, pe baza unor subperioade neegale între ele, se va recurge la formula mediei geometrice ponderate (19.30). Pentru exemplificare, calculăm ritmul mediu anual de variație al perioadei 1938-1976 pe baza ritmurilor medii de variație ale perioadelor 1938-1950, 1950-1960 și 1960-1976:

$$\bar{R}_{1938-1950} = \sqrt[12]{1,8699} = 1,053$$

$$\bar{R}_{1950-1960} = \sqrt[10]{3,6204} = 1,137$$

$$\bar{R}_{1960-1976} = \sqrt[16]{7,6165} = 1,135$$

iar

$$\bar{R}_{1938-1976} = \sqrt[38]{1,053^{12} \cdot 1,137^{10} \cdot 1,135^{16}} = 1,109$$

Formula (19.30) permite atît determinarea ritmului mediu de creștere pentru întreaga perioadă, pe baza ritmurilor medii de creștere ale diferitelor subperioade, cît și determinarea ritmului mediu al unei subperioade atunci cînd se cunosc ritmurile medii de creștere ale celorlalte subperioade și ritmul mediu de creștere pentru întreaga perioadă.

Pe baza ritmurilor medii ale variației se pot determina ritmurile medii ale creșterii sau ale sporului cu ajutorul formulei:

$$\bar{r}_y = \bar{R}_y - 1 \quad \text{sau} \quad \bar{r}_y\,\% = \bar{R}_y\,\% - 100$$

Exprimat în procente ritmul mediu al creșterii arată cu cît a crescut în medie în fiecare an al perioadei luate în calcul, nivelul unui anumit fenomen.

Metoda mediei geometrice de determinare a ritmului mediu de dezvoltare a fenomenelor, prezintă dezavantajul că ia în calcul numai valorile extreme ale unei serii y_0 şi y_n şi neglijează toţi ceilalţi termeni intermediari.

Din această cauză metoda mediei geometrice se utilizează în practică pentru determinarea ritmului mediu de dezvoltare a fenomenelor numai pentru acele serii dinamice în care ritmurile de variaţiei cu bază mobilă nu diferă în mod substanţial de la o perioadă la alta.

Pentru înlăturarea acestor deficienţe, în literatura de specialitate s-au propus şi alte metode, printre care, un loc mai important îl ocupă metoda mediei parabolice şi metoda trendului exponenţial. Aplicarea uneia sau alteia dintre aceste metode se face ţinînd seama de proprietăţile lor, de posibilităţile de obţinere a datelor ca şi de esenţa fenomenului studiat.

19.3.1.2. Metoda mediei parabolice

Metoda mediei parabolice de calculare a ritmului mediu înlătură parţial dezavantajele metodei mediei geometrice, luînd în calcul suma valorilor termenilor seriei şi valoarea termenului iniţial. In cazul în care termenii unei serii dinamice variază (aproximativ) în progresie geometrică, are loc relaţia:

$$y_0 + y_1 + y_2 + \dots + y_i \dots + y_n = \sum_0^n y_i$$

sau

$$y_0\bar{R}^0 + y_0\bar{R} + y_0\bar{R}^2 + \dots + y_0\bar{R}^i + \dots + y_0\bar{R}^n = \sum_0^n y_i \qquad (19.20)$$

$$y_0(1 + \bar{R} + \bar{R}^2 + \dots + \bar{R}^n) = \sum_0^n y_i$$

$$1 + \bar{R} + \bar{R}^2 + \dots + \bar{R}^i + \dots + \bar{R}^n = \frac{\sum_0^n y_i}{y_0} \qquad (19.21)$$

Primul membru al relaţiei (19.21) reprezintă suma termenilor unei progresii geometrice:

$$1 + \bar{R} + \bar{R}^2 + \ldots + \bar{R}^i + \ldots + \bar{R}^n = \frac{\bar{R}^{n+1} - 1}{\bar{R} - 1}$$

deci

$$\frac{\bar{R}^{n+1} - 1}{\bar{R} - 1} = \frac{\sum_0^n y_i}{y_0} \qquad (19.22)$$

Determinarea ritmului mediu pe baza relaţiei (19.22) se face prin încercări succesive şi de aceea necesită un volum mare de calcule.

In practică, pentru înlăturarea acestei deficienţe, statisticianul chinez Fin Di Şi a întocmit un tabel ajutător, pe baza căruia se poate determina ritmul mediu, cu ajutorul metodei mediei parabolice, pentru diferite perioade de timp.

Formula (19.22) scoate în evidenţă faptul că metoda mediei parabolice de determinare a ritmului mediu ţinea seama de primul termen al seriei dinamice, pentru care se calculează, şi de suma tuturor termenilor ei. Exagerarea importanţei primului termen al seriei reprezintă de fapt şi principala deficienţă a acestei metode.

19.3.1.3. Metoda trendului exponenţial

Deficienţele metodei mediei geometrice şi a metodei mediei parabolice de calcul al ritmului mediu se înlătură, în mare parte, prin utilizarea metodei trendului exponenţial recomandată de literatura de specialitate.

Dacă ritmurile variaţiei cu bază mobilă sînt apropiate ca mărime, atunci variaţia unui fenomen se poate descrie cu ajutorul unei funcţii exponenţiale:

$$y_n = ab^n \qquad (19.23)$$

Viteza relativă de creştere a fenomenului y este tocmai parametrul b care joacă deci rolul de ritm mediu de creştere.

Substituind în relaţia (19.23) pe b cu $\bar{R}$ rezultă:

$$y_n = a\,\bar{R}^n \qquad (19.24)$$

Parametrii a şi $\bar{R}$ ai funcţiei (19.24) se pot determina uşor cu ajutorul metodei celor mai mici pătrate. Prin logaritmare, relaţia (19.24) devine:

$$\log y_n = \log a + n \log \bar{R} \qquad (19.25)$$

Sistemul de ecuaţii normale, ce se obţine în urma aplicării metodei celor mai mici pătrate, ia forma:

$$\begin{aligned} n \log a + \log \bar{R} \sum n_i &= \sum \log y_i \\ \log a \sum n_i + \log \bar{R} \sum n_i^2 &= \sum n_i \log y_i \end{aligned} \qquad (19.26)$$

Deoarece n_i reprezintă timpul, care în mod obişnuit în cadrul seriilor dinamice variază în progresie aritmetică, sistemul (19.26) se poate aduce la o formă mai simplă pe baza substituţiei $\sum n_i = 0$, astfel:

$$\begin{aligned} n \log a &= \sum \log y_i \\ \log \bar{R} \sum n_i^2 &= \sum n_i \log y_i \end{aligned} \qquad (19.27)$$

Din ecuaţia a doua a sistemului (19.27) rezultă imediat expresia ritmului mediu:

$$\log \bar{R} = \frac{\sum n_i \log y_i}{\sum n_i^2} \qquad (19.28)$$

iar

$$\bar{R} = \text{antilog} \frac{\sum n_i \log y_i}{\sum n_i^2} \qquad (19.29)$$

Rezultă că, atunci cînd se cunosc doar termenii extremi ai unei perioade se utilizează, pentru calculul ritmului mediu, metoda mediei geometrice, iar dacă se cunosc şi termenii intermediari, este recomandabil să se utilizeze metoda mediei parabolice sau metoda trendului exponenţial.

19.3.2. Determinarea ritmului mediu pentru fenomenele complexe

Metodele de calculare a ritmului mediu de creştere prezentate pînă acum sînt aplicabile în cazul unor fenomene generale sau singulare, deoarece fac abstracţie de structura lor.

Luînd în calcul structura fenomenelor, se pot distinge fenomene complexe care iau naştere prin însumarea unor fenomene simple, iar altele ca rezultat al produsului mai multor fenomene simple. Astfel, fenomenele complexe pot fi descrise cu ajutorul unor funcţii compuse de forma:

$$Y = f(X_0, X_1, X_2, \ldots, X_k) = x_0(t) + x_1(t) + x_2(t) + \ldots + x_i(t) + \ldots + x_k(t) = \sum_0^k x_i(t) \qquad (19.31)$$

$$Y = f(X_0, X_1, X_2, \ldots, X_k) = x_0(t) \cdot x_1(t) \cdot x_2(t) \ldots x_i(t) \ldots x_k(t) = \prod_0^k x_i(t) \qquad (19.32)$$

De exemplu, producţia globală lunară (Q) este egală cu produsul dintre numărul mediu scriptic lunar al personalului (T) şi productivitatea lunară a muncii (Wl), adică:

$$Q = T \cdot Wl$$

La rîndul ei, productivitatea lunară a muncii se poate descompune în alţi doi factori şi anume, se poate scrie ca produs între durata medie a lunii de lucru în zile (Dz) şi productivitatea zilnică (Wz), astfel:

$$Wl = Dz \cdot Wz$$

iar

$$Q = T \cdot Dz \cdot Wz$$

Productivitatea zilnică a muncii se poate descompune în alţi doi factori şi anume durata medie a zilei de lucru, exprimată în ore, (Dh) şi productivitatea orară a muncii (Wh):

$$Wz = Dh \cdot Wh$$

şi

$$Wl = Dz \cdot Dh \cdot Wh$$

iar $Q = T.Ds.Dh.Wh$

S-a descompus astfel fenomenul complex - producţia globală lunară - în produsul a patru factori:

- numărul personalului;
- durata medie a lunii de lucru în zile;
- durata medie a zilei de lucru în ore;
- productivitatea orară a muncii.

Un alt exeplu, volumul total al desfacerilor de mărfuri prin comerţul socialist este format din suma dintre volumul mărfurilor alimentare, volumul mărfurilor nealimentare şi volumul alimentaţiei publice. Volumul consumului de materii şi materiale necesare pentru producerea unui produs, este format din suma tuturor felurilor de materii şi materiale ce se consumă.

In cazul fenomenelor complexe, care reprezintă rezultatul însumării mai multor fenomene simple, ritmul mediu al fenomenului complex se va calcula după formula generală (19.18), în care:

$$y_n = x_0(n)+x_1(n)+x_2(n)+\ldots+x_i(n)+\ldots+x_k(n)$$

iar

$$y_0 = x_0(o)+x_1(o)+x_2(o)+\ldots+x_i(o)+\ldots+x_k(o)$$

$$t = 0,1,2,\ldots,n$$

deci

$$\bar{R}_y = \sqrt[n]{\frac{y_n}{y_0}} = \sqrt[n]{\frac{\sum x_i(n)}{\sum x_i(o)}} \qquad (19.33)$$

Pentru exemplificare, considerăm datele din tabelul 19.5.

Tabelul 19.5. Volumul şi structura vînzărilor de mărfuri cu amănuntul în RSR prin comerţul socialist în anii 1970 şi 1976 (în preţurile fiecărui an) - milioane lei

Anii	Total (milioane lei)	din care:		
		Mărfuri alimentare	Alimentaţia publică	Mărfuri nealimentare
1970	96708	34633	12899	49176
1976	159446	57638	22008	79800

Sursa: Anuarul statistic al R.S.R., 1977, D.C.S., p. 404.

Aplicînd formula (19.33) pentru volumul total al vînzărilor de mărfuri cu amănuntul, rezultă că ritmul mediu anual în perioada 1970-1976 a fost:

$$\bar{R} = \sqrt[6]{\frac{159446}{96708}} = 1,087$$

- pentru mărfuri alimentare: $\bar{R}_a = \sqrt[6]{\frac{57638}{34633}} = 1,08$

- pentru alimentaţia publică: $\bar{R}_{ap} = \sqrt[6]{\frac{22008}{12899}} = 1,083$

- pentru mărfuri nealimentare: $\bar{R}_{na} = \sqrt[6]{\frac{79800}{49176}} = 1,083$

Dar elementele formulei (19.33) se mai pot scrie:

$$x_0(n) = x_0(o).\bar{R}^n_{x_0}$$

$$x_1(n) = x_1(o).\bar{R}^n_{x_1}$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$x_i(n) = x_i(o).\bar{R}^n_{x_i}$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$x_k(n) = x_k(o).\bar{R}^n_{x_k}$$

Însumînd membru cu membru, rezultă:

$$\sum_0^k x_i(n) = \sum_0^k x_i(o)\,\bar{R}^n_{x_i} \qquad (19.34)$$

Pe baza relaţiei (19.34), formula ritmului mediu (19.33) devine:

$$\bar{R} = \sqrt[n]{\frac{\sum x_i(o)\,\bar{R}^n_{x_i}}{\sum x_i(o)}} \qquad (19.35)$$

Din relaţia (19.35) rezultă posibilitatea calculării ritmului mediu al unui fenomen complex, pentru o anumită perioadă de

timp n , atunci cînd se cunosc ritmurile medii ale fenomenelor simple ($\bar{R}_{x_i}$) şi nivelurile acestora din perioada de bază $x_i(o)$.

Ritmul mediu anual de creştere al volumului total al vînzărilor de mărfuri cu amănuntul se poate determina şi cu ajutorul formulei (19.35):

$$\bar{R} = \sqrt[6]{\frac{34633.1,080^6+12899.1,083^6+49176.1,083^6}{34633 + 12899 + 49176}} = 1,087$$

Dacă se substituie y_n şi y_o cu expresiile lor, în formula (19.33), rezultă:

$$\bar{R}_y = \sqrt[n]{\frac{\sum x_i(n)}{\sum x_i(o)}} \qquad (19.36)$$

Relaţia (19.36) dă posibilitatea determinării oricărui ritm mediu al unui fenomen simplu atunci cînd se cunoaşte ritmul mediu al fenomenului complex pentru aceeaşi perioadă de timp şi ritmurile medii ale celorlalte fenomene simple. Astfel, să presupunem că se cunosc: ritmul mediu al fenomenului complex $\bar{R}_y$,ritmurile medii medii ale tuturor fenomenelor simple ce intră în componenţa lui y , cu excepţia lui X_o , nivelurile acestor fenomene din perioada de bază $x_i(o)$. Se cere să se determine ritmul mediu al fenomenului X_o, pentru aceeaşi perioadă. Formula generală a ritmului mediu al fenomenului X_o va fi:

$$\bar{R}_{x_o} = \sqrt[n]{\frac{x_o(n)}{x_o(o)}} \qquad (19.37)$$

Din această formulă nu se cunoaşte elementul $x_o(n)$, a cărui expresie se poate determina după formula:

$$x_o(n) = x_o(o).\hat{R}^n_{x_o} = y_n - \sum_1^k x_i(n) = y_o.\hat{R}^n_y - \sum_1^k x_i(o)\bar{R}^n_{x_i} \qquad (19.38)$$

Substituind pe (19.38) în relaţia (19.37), rezultă:

$$\bar{R}_{x_0} = \sqrt[n]{\frac{y_0 \bar{R}_y^n - \sum_1^k x_i(o)\, \bar{R}_{x_i}^n}{x_0(o)}} \qquad (19.39)$$

sau

$$\bar{R}_{x_0} = \sqrt[n]{\frac{y_n - \sum_1^k x_i(n)}{x_0}}$$

In cazul fenomenelor complexe de tipul (19.32) ritmul mediu de creştere se determină cu ajutorul formulei:

$$\bar{R}_y = \sqrt[n]{\frac{y_n}{y_0}} = \sqrt[n]{\frac{\prod_1^k x_i(n)}{\prod_1^k x_i(o)}} \qquad i = 0,1,2,\ldots,k \qquad (19.40)$$

Dezvoltînd produsele şi grupînd factorii, în ultima expresie a relaţiei (19.40), rezultă :

$$\bar{R}_y = \sqrt[n]{\frac{x_0(n)}{x_0(o)} \cdot \frac{x_1(n)}{x_0(o)} \cdot \frac{x_2(n)}{x_2(o)} \cdots \frac{x_i(n)}{x_i(o)} \cdots \frac{x_k(n)}{x_k(o)}} \qquad (19.41)$$

dar

$$\frac{x_i(n)}{x_i(o)} = R_{x_i} \qquad i = 0,1,2,\ldots,k \qquad (19.42)$$

R_{x_i} = ritmul variaţiei cu bază fixă pe întreaga perioadă a fenomenului x_i .

Substituind în relaţia (19.41) pe (1942), se obţine:

$$\bar{R}_y = \sqrt[n]{R_{x_0} \cdot R_{x_1} \cdot R_{x_2} \cdots R_{x_i} \cdots R_{x_k}} \qquad (19.43)$$

sau mai sintetic:

$$\bar{R}_y = \sqrt[n]{\prod_1^k R_{x_i}} \qquad i = 0,1,2,\ldots,k \qquad (19.44)$$

Intre nivelurile inițiale și nivelurile finale ale diferiteler fenomene singulare există relațiile:

$$\begin{aligned} x_0(n) &= x_0(o).\bar{R}^n_{x_0} \\ x_1(n) &= x_1(o).\bar{R}^n_{x_1} \\ x_2(n) &= x_2(o).\bar{R}^n_{x_2} \\ &\cdots\cdots\cdots\cdots \\ x_i(n) &= x_i(o).\bar{R}^n_{x_i} \\ &\cdots\cdots\cdots\cdots \\ x_k(n) &= x_k(o).\bar{R}^n_{x_k} \end{aligned} \qquad (19.45)$$

Ținînd seama de relațiile (19.45) , ritmul mediu al fenomenului complex $\bar{R}_y$, devine:

$$\bar{R}_y = \sqrt[n]{\frac{\prod_o^k x_i(o)\ \bar{R}^n_{x_i}}{\prod_o^k x_i(o)}} \qquad (19.46)$$

sau

$$\bar{R}_y = \sqrt[n]{\prod_o^k \bar{R}^n_{x_i}} \qquad (19.47)$$

$$\bar{R}_y = \prod_o^k \bar{R}_{x_i} \qquad (19.48)$$

Considerăm, pentru exemplificare, datele din tabelul 19.6.

Tabelul 19.6. Populația și venitul național al României în anii 1970 și 1976

	U.M.	1970	1976
(Populația (la 1 iulie)	mii locuitori	20252,5	21445,7
Venitul național (prețuri comparabile 1963)	miliarde lei	212,1	400,0

Sursa: Anuarul Statistic al R.R.S.,1977,DCS,p.84

Venitul naţional mediu pe un locuitor a fost:
- în anul 1970 de 10 473 lei
- în anul 1976 de 18 651 lei

Ritmul mediu anual de creştere a populaţiei României în perioada 1971-1976 a fost:

$$\bar{R}_p = \sqrt[6]{\frac{21445,7}{20252,5}} = 1,01$$

Ritmul mediu anual de creştere a venitului naţional pe un locuitor în perioada 1971-1976 a fost:

$$\bar{R}_v = \sqrt[6]{\frac{18651}{10473}} = 110,1$$

Ritmul mediu anual de creştere a venitului naţional total în perioada 1971-1976 a fost:

$$\bar{R}_V = \sqrt[6]{\frac{400,0}{212,1}} = 111,1$$

Intrucît venitul naţional total se poate obţine prin înmulţirea numărului populaţiei cu venitul naţional pe un locuitor, rezultă, conform relaţiei (19.48) că:

$$\bar{R}_V = \bar{R}_p \cdot \bar{R}_v = 1,01 \cdot 110,1 \simeq 111,1$$

Relaţia (19.48) scoate în evidenţă legătura dintre ritmul mediu al fenomenului complex şi ritmurile medii ale fenomenelor simple: ritmul mediu al fenomenului complex este egal cu produsul ritmurilor medii ale fenomenelor simple de tipul (19.32).

19.4. Prognoza fenomenelor pe baza indicatorilor mediilor de creştere

Prima problemă pe care trebuie să şi-o pună orice cercetător asupra evoluţiei viitoare a unui fenomen este stabilirea pe cale logică a existenţei viitoare a acestuia. Altfel spus, cercetătorul trebuie să stabilească dacă prognoza unui fenomen, într-un orizont dat, este sau nu posibilă. Prognoza unui fenomen poate fi

întotdeauna efectuată cînd există o continuitate în evoluţia lui, cînd există deci o legătură între nivelul prezent şi cel viitor, cînd evoluţia din trecut determină într-o anumită măsură, evoluţia lui prezentă şi viitoare, cînd în evoluţia unui fenomen putem să descoperim o anumită tendinţă de desvoltare. Dimpotrivă, dacă situaţia unui fenomen s-ar prezenta astfel încît evoluţia lui viitoare ar fi în totalitate ruptă, independentă de prezent şi trecut, să nu decurgă din ceea ce este şi din ceea ce a fost, atunci prognoza lui nu ar fi posibilă. Viitorul fenomenelor sociale şi economice se bazează pe modelul lor de evoluţie şi, uneori, pe ipoteza că în decursul orizontului de prognoză considerat, aceste fenomene nu vor fi supuse unor modificări esenţiale.

Două dintre modelele statistice de prognoză, cele mai vechi, mai simple şi mai des utiliuate în practică, se fundamentează pe ipoteza variaţiei fenomenelor în progresie aritmetică şi în progresie geometrică.

Ţinînd seama de aceste două fenomene se disting două categorii de metode de prognoză a evoluţiei fenomenelor sociale şi economice:

- metode de prognoză bazate pe sporul mediu;
- metode de prognoză bazate pe ritmul mediu de creştere.

19.4.1. Metode de prognoză bazate pe sporul mediu

Evoluţia unui fenomen în progresie aritmetică se realizează atunci cînd sporurile anuale cu bază mobilă sînt apropiate ca mărime, iar forma evoluţiei este liniară.

In condiţiile evoluţiei uniforme, liniare a unui fenomen, după modelul unei progresii aritmetice, se pot rezolva mai multe probleme de prognoză:

1) <u>Determinarea nivelului unui fenomen peste t ani</u>

Să presupunem că un anumit fenomen y a înregistrat, într-o perioadă anterioară, o creştere absolută relativ constantă, dată de un spor mediu egal cu Δ_y. Se cere să se determine care va fi nivelul fenomenului y peste n ani.

In ipoteza că şi în perspectivă fenomenul y se va de-

volta tot în progresie aritmetică, cu aceeaşi raţie ($\bar{\Delta}_y$), valorile pe care le va atinge în următorii ani sînt:

$$\begin{aligned} y_0 &= y_0 \\ y_1 &= y_0 + \bar{\Delta}_y \\ y_2 &= y_0 + 2\bar{\Delta}_y \\ &\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots \\ y_n &= y_0 + n\bar{\Delta}_y \end{aligned} \qquad (19.49)$$

De exemplu, în perioada 1970-1976 producţia de energie electrică a României a crescut cu un spor mediu anual egal cu $\bar{\Delta}_y$ = 3863 milioane kwh (vezi tabelul 19.2). In ipoteza că în viitor producţia de energie electrică va creşte în condiţii similare, ea va ajunge în 1980 la 89170 milioane kwh şi la 108485 milioane kwh în anul 1985.

- 2) <u>Stabilirea orizontului prognozei în care un anumit fenomen va atinge un nivel prestabilit</u>

Să considerăm că, pentru a ajunge la îndeplinirea unor programe de dezvoltare economică şi socială a unei zone într-o ţară, este necesar ca să se dubleze volumul producţiei unui anumit produs de bază y. Cunoscînd sporul mediu anual de creştere al fenomenului y, se cere să se determine perioada de timp necesară pentru dublarea fenomenului considerat. Calculele de prognoză se vor efectua în ipoteza că şi în viitor fenomenul y va evolua după aceeaşi tendinţă centrală, dată de modelul progresiei aritmetice, observată în perioada anterioară.

Dacă notăm:

y_0 - nivelul fenomenului y într-un moment anterior celui de calcul;

y_k - nivelul fenomenului y în momentul efectuării calculelor;

y_n - nivelul fenomenului y în momentul dublării;

$\bar{\Delta}_y$ - sporul mediu al fenomenului y observat în perioada anterioară,

atunci evoluţia viitoare a lui y va fi descrisă de relaţia :

$$y_n = y_k + (n-k)\,\bar{\Delta}_y \qquad (19.50)$$

Substituind pe $y_n = 2\,y_k$ şi $n - k = t$, în relaţia (19.50), rezultă:

$$2\,y_k = y_k + t\,\bar{\Delta}_y$$

$$y_k = t\,\bar{\Delta}_y$$

de unde:

$$t = \frac{y_k}{\bar{\Delta}_y} \qquad (19.51)$$

Din relaţia (19.51) rezultă că timpul necesar dublării nivelului unui anumit fenomen este egal cu raportul dintre nivelul prezent al fenomenului respectiv şi sporul lui mediu anual.

Reluînd exemplul anterior, rezultă că producţia de energie electrică a României s-ar dubla, în baza formulei (19.51), într-o perioadă de timp egală cu 15 ani, adică în anul 1991:

$$t = \frac{58266}{3863} = 15 \text{ ani}$$

Dublarea producţiei actuale de energie electrică se va produce după 15 ani, dacă şi numai dacă în cadrul acestei perioade viitoare ea va creşte în medie cu acelaşi spor mediu anual.

3) Stabilirea perioadei de timp t necesară pentru egalizarea nivelului de dezvoltare a două zone A şi B

Să presupunem că o zonă (localitate, judeţ, regiune etc.) A îşi propune să ajungă din urmă o altă zonă B mai dezvoltată din punct de vedere al unui anumit indicator y . Dintr-o perioadă anterioară se cunoaşte sporul mediu anual de creştere a lui y, în fiecare zonă, respectiv $\bar{\Delta}_y^A$ şi $\bar{\Delta}_y^B$ şi se cere să se determine peste cîţi ani "t" zona A va ajunge din urmă zona B . La data efectuării prognozei nivelul indicatorului y în cele două zone era y_o^A şi y_o^B .

Această problemă are sens numai dacă dezvoltarea prezentă

şi viitoare a zonei A este şi va fi mai rapidă decît a zonei B.

In momentul egalizării nivelului lui y , în cele două zone, va fi:

$$y_t^A = y_o^A + t\ \bar{\Delta}_t^A$$

şi

$$y_t^B = y_o^B + t\ \bar{\Delta}_y^B$$

deci

$$y_o^A + t\bar{\Delta}_y^A = y_o^B + t\bar{\Delta}_y^B$$

de unde

$$t\ (\bar{\Delta}_y^A - \bar{\Delta}_y^B) = y_o^B - y_o^A$$

iar

$$t = \frac{y_o^B - y_o^A}{\bar{\Delta}_y^A - \bar{\Delta}_y^B} \qquad (19.52)$$

Aceasta înseamnă că perioada de timp necesară egalizării nivelului de dezvoltare al celor două zone,este egală cu raportul dintre diferenţa între nivelurile de dezvoltare ale zonelor în momentul calculului şi diferenţa sporurilor medii ale celor două fenomene obţinute pentru perioada anterioară.

Să considerăm, pentru exemplificare, următoarea problemă: Cu privire la producţia de ciment a României şi Cehoslovaciei se cunosc datele din tabelul 19.7.

Tabelul 19.7. Producţia de ciment a României şi Cehoslovaciei în anii 1965 - 1975

Ţara	Ciment mii tone		Populaţia (milioane locuitori	
	1965	1975	1965	1975
R.S.Cehoslovacia	5713	9305	14,16	14,80
R.S. România	5406	11520	19,03	21,25

Sursa: Anuarul statistic al R.S.R. 1977, D.C.S.,Bucureşti, p. 538-540, 568.

Scanned with OKEN Scanner

Tabelul 19.8. Producţia de ciment pe un locuitor a României şi Cehoslovaciei (în kg)

Ţara \ Anul	1965	1975
R.S.Cehoslovacia	403	629
R.S.România	284	542

In perioada 1965-75 sporul mediu al producţiei de ciment pe locuitor a fost:

- în Cehoslovacia: $\bar{\Delta}_C = \frac{629-403}{5} = \frac{226}{5} = 45$ mii tone
- în România : $\bar{\Delta}_R = \frac{542-284}{5} = \frac{258}{5} = 52$ mii tone

Decalajul, care mai separă România de Cehoslovacia în privinţa producţiei de ciment pe un locuitor, va dispare peste o perioadă de timp dată de formula (19.52):

$$t = \frac{629 - 542}{52 - 45} = \frac{87}{7} = 12 \text{ ani}$$

Dispariţia decalajului va fi posibilă peste 12 ani numai dacă şi în această perioadă producţia de ciment a celor două ţări va cunoaşte acelaşi spor mediu ca şi în perioada 1965-1975.

4) <u>Egalizarea nivelului de dezvoltare a două zone peste o perioadă de timp prestabilă</u>

Să presupunem că o zonă A, mai puţin dezvoltată, îşi propune să ajungă din urmă o zonă B, mai dezvoltată în privinţa indicatorului y într-o anumit perioadă de timp prestabilită t. Se cere să se determine sporul mediu anual de creştere al indicatorului y în zonă A($\bar{\Delta}_y^A$) necesar pentru asigurarea egalizării cu zona B într-un timp t.

Ca şi în cazul anterior, în momentul egalizării celor două zone se va verifica cu relaţia:

$$y_0^A + \bar{\Delta}_y^A = y_0^B + t\,\bar{\Delta}_y^B$$

Presupunînd că zona B îşi va păstra în viitor acelaşi spor mediu anual de creştere, atunci din această relaţie va trebui estimat Δ_y^A, ceilalţi termeni fiind cunoscuţi:

$$t\bar{\Delta}_y^A = y_0^B - y_0^A + t\bar{\Delta}_y^B$$

de unde

$$\bar{\Delta}_y^A = \frac{y_0^B - y_0^A}{t} + \bar{\Delta}_y^B \qquad (19.53)$$

Sublinem faptul că toate prognozele fenomenelor cu ajutorul metodei sporului mediu se bazează pe ipoteza că şi în perspectivă fenomenele luate în calcul se vor dezvolta tot în progresie aritmetică.

Considerînd datele din problema precedentă şi presupunînd că România îşi propune să înlăture decalajul ce o separă în prezent (1976) de Cehoslovacia în numai 5 ani. Pentru a-şi îndeplini acest obiectiv va fi necesar ca producţia de ciment să crească în fiecare an cu un spor mediu anual egal cu:

$$\bar{\Delta}_y^R = \frac{629 - 542}{5} + 45 = 62{,}4 \text{ mii tone}$$

Aceasta înseamnă că, în condiţiile date, pentru înlăturarea decalajului dintre România şi Cehoslovacia în privinţa producţiei de ciment pe locuitor în decurs de 5 ani, va fi necesar ca producţia României în următorii cinci ani să crească în medie, în fiecare an cu 62,4 mii tone.

19.4.2. Metode de prognoză bazate pe ritmul mediu

In ipoteza evoluţiei unui fenomen după modelul unei progresii geometrice, prognoza lui se poate face cu ajutorul unor metode bazate pe ritmul mediu de creştere, care reprezintă de fapt o estimare a raţiei acestei progresii.

Evoluţia unui fenomen după modelul unei progresii geometrice are loc cînd ritmurile de variaţie cu bază mobilă sînt apropiate ca mărime.

1. Stabilirea nivelului de dezvoltare al unui anumit fenomen peste t ani.

Considerînd că fenomenul y s-a dezvoltat în perioada anterioară cu un ritm mediu anual de creştere $\bar{R}_y$ şi în momentul efectuării prognozei y_0, se cere să se determine care va fi nivelul acestui fenomen peste t ani.

Creşterea lui y în progresie geometrică se exprimă astfel:

$$y_0 = y_0$$

$$y_1 = y_0 \bar{R}_y$$

$$y_2 = y_0 \bar{R}_y^2$$

$$\cdots\cdots\cdots$$

$$y_n = y_0 \bar{R}_y^n \qquad (19.54)$$

Considerăm din nou exemplul din tabelul 19.3 în condiţiile ritmului mediu anual de creştere a producţiei de energie electrică realizat în perioada 1965-1976 de 1,117, în anul 1980 se va ajunge la o producţie egală cu:

$$y_{1980} = y_{1976} \cdot \bar{R}^4 = 58266.1,117^4 = 90604 \text{milioane kwh}$$

Deci, evoluţia fenomenului y în perspectivă este descrisă de formula (19.54), care reprezintă termenul general al unei progresii geometrice cu raţia egală cu ritmul mediu de creştere.

2. Stabilirea orizontului prognozei în care un anumit fenomen va atinge un nivel prestabilit.

Să presupunem că se cere să se determine perioada de timp necesară dublării fenomenului y. In condiţiile notaţiilor anterioare, în momentul dublării fenomenului y, se va verifica relaţia:

$$y_t = y_0 \bar{R}_y^t$$

dar

$$y_t = 2 y_0$$

deci

$$2 = \bar{R}_y^t$$

şi

$$\log 2 = t \log \bar{R}_y$$

de unde

$$t = \frac{\log 2}{\log \bar{R}_y} \qquad (19.55)$$

In condiţiile datelor din problema anterioară, producţia de energie electrică se va dubla într-un interval de timp egal cu 6 ani:

$$t = \frac{0,30103}{0,04805} = 6,26 \text{ ani}$$

3. Stabilirea perioadei de timp necesare pentru egalizarea nivelului de dezvoltare a două zone.

Să presupunem că o zonă A , mai puţin dezvoltată în privinţa indicatorului y , îşi presupune în viitor să ajungă din urmă o zonă B mai dezvoltată. Pe baza analizei evoluţiei celor două zone, în trecut, se cunosc ritmurile lor medii de dezvoltare respectiv $\bar{R}_y^B$ şi nivelurile fenomenului considerat din momentul prognozei era y_o^A şi respectiv y_o^B .

Se cere să se determine, în condiţiile evoluţiei celor două fenomene şi în viitor în progresie geometrică cu aceleaşi raţii, peste cîţi ani t zona A va ajunge din urmă zona B .

Problema are sens numai dacă ritmul de creştere în zona A este mai mare decît ritmul de creştere din zona B .

Pe baza formulei (19.54) în momentul egalizării va avea loc relaţia :

$$y_o^A \cdot \bar{R}_A^t = y_o^B \bar{R}_B^t$$

Din această relaţie trebuie determinat t :

$$\log y_o^A + t \log \bar{R}_A = \log y_o^B + t \log \bar{R}_B$$

iar

$$t (\log \bar{R}_A - \log \bar{R}_B) = \log y_o^B - \log y_o^A$$

de unde

$$t = \frac{\log y_o^B - \log y_o^A}{\log \bar{R}_A - \log \bar{R}_B} \qquad (19.56)$$

Pentru exemplificare considerăm datele din tabelul 19.7. Ritmul mediu anual de variaţie a producţiei de ciment în perioada 1965-1975 a fost:

- în România : $\bar{R}_R = \frac{542}{284} = 1,067$ ori

- în Cehoslovacia: $\bar{R}_C = \frac{629}{403} = 1,045$ ori

In aceste condiţii egalizarea nivelului producţiei de ciment pe locuitor se va produce peste 7,3 ani:

$$t = \frac{\log 629 - \log 542}{\log 1,067 - \log 1,045} = \frac{2,79865 - 2,73400}{0,02816 - 0,01912} = 7,3 \text{ ani}$$

4. Egalizarea nivelului de dezvoltare a două zone într-o perioadă de timp prestabilită.

Să considerăm că o zonă A îşi propune să ajungă din urmă în privinţa indicatorului y , o zonă B mai dezvoltată, într-o perioadă de timp dinainte stabilită t . In ipoteza că zona B îşi va păstra acelaşi ritm mediu anual de creştere $\bar{R}_B$ - estimat din evoluţia trecută - se cere să se determine cu ce ritm mediu anual trebuie să se dezvolte zona A în aşa fel încît după t ani să se producă egalizarea între ele.

In momentul egalizării se va verifica relaţia:

$$y_o^A \cdot \bar{R}_A^t = y_o^B \cdot \bar{R}_B^t$$

Din această relaţie trebuie determinat $\bar{R}_A$:

$$\log y_o^A + t \log \bar{R}_A = \log y_o^B + t \log \bar{R}_B$$

$$t \log \bar{R}_A = \log y_o^B - \log y_o^A + t \log \bar{R}_B$$

de unde

$$\log \bar{R}_A = \frac{\log y_o^B - \log y_o^A}{t} + \log \bar{R}_B \qquad (19.57)$$

In condiţiile datelor din problema anterioară, să presupunem că România îşi propune să înlăture decalajul care o separă de Cehoslovacia, în privinţa producţiei de ciment pe locuitor,

în decurs de 5 ani şi se cere să se determine cu ce ritm mediu anual trebuie să se dezvolte în decursul acestei perioade pentru îndeplinirea obiectivului propus. Aplicînd formula (19.57) rezultă:

$$\bar{R}_R = \text{antilog}\left(\frac{\log y_n^B - \log y_n^A}{t} + \log \bar{R}_C\right) =$$

$$= \text{anitolg}\left(\frac{2{,}79865 - 2{,}73400}{5} + 0{,}01912\right) = 1{,}077 \text{ ori}$$

Metodele de prognoză bazate pe sporul mediu şi pe ritmul mediu de creştere au o largă utilizare în practică, datorită operativităţii rezultatelor, simplităţii calculelor şi analizei rezultatelor ca şi gradului de precizie satisfăcător. Aceste metode se bazează pe unele generalizări cu caracter de legitate de dezvoltare, obţinute pe baza unor analize multilaterale a fenomenelor economice şi sociale cercetate şi se utilizează în practică pentru fenomene omogene a căror structură prezintă o anumită stabilitate în timp şi spaţiu.

20. TENDINŢA GENERALĂ DE EVOLUŢIE A FENOMENELOR ÎN TIMP

20.1. Noţiunea de tendinţă generală de dezvoltare

Evoluţia fenomenelor în timp este supusă acţiunii unui complex de factori în rîndul cărora se pot evidenţia cîteva categorii mai importante şi anume: <u>factori esenţiali</u>, care ţin de natura internă de dezvoltare, <u>factori întîmplători</u>, accidentali, <u>factori cu acţiune sezonieră</u> şi <u>factori cu acţiune ciclică</u>.

Factorii esenţiali determină <u>tendinţa generală de dezvoltare</u> a fenomenelor - <u>legitatea de dezvoltare</u> - , stabilă şi, de regulă, de lungă durată numită şi <u>trend</u>. Separarea componentei de bază a dezvoltării fenomenelor în timp - trendul - se obţine prin operaţia de ajustare - numită uneori şi netezire - a seriilor dinamice.

Determinarea tendinţei generale de dezvoltare a fenomenelor prezintă o importanţă deosebită în cunoaşterea legităţilor lor de dezvoltare, iar aceasta constituie baza de fundamentare ştiinţifică a deciziilor privind evoluţia viitoare a fenomenelor şi proceselor din natură şi societate.

Descoperirea legităţilor de dezvoltare în timp ale fenomenelor nu este posibilă decît prin analiza acestora pe o perioadă suficient de lungă din trecut, aceasta reprezentînd o cerinţă esenţială a acţiunii legii numerelor mari.

A ajusta o serie dinamică (t_i, y_i) înseamnă a înlocui termenii seriei reale cu alţi termeni obţinuţi cu ajutorul unor metode diferite, şi care reflectă tendinţa generală de lungă durată de dezvoltare a fenomenelor. Termenii seriei ajustate corespund valorilor succesive ale unei funcţii continue $y(t)$.

Timpul joacă rolul de variabilă independentă sau explicativă, iar seria dinamică se poate considera ca o funcţie de timpul "t"

$$y_t = f(t) + \varepsilon t \qquad (20.1)$$

unde εt reprezintă o parte din indicatorul y_t care nu poate fi explicată pe seama funcţiei $f(t)$ şi se numeşte <u>reziduu</u> sau <u>fluctuaţie aleatoare</u>. Mărimea şi sensul termenului εt variază în timp odată cu indicatorul y şi depinde numai de întîmplare. Suma tuturor valorilor εt, pe întreaga scară de variaţie a timpului, tinde către zero pe măsură ce numărul cazurilor observate (lungimea seriei dinamice) creşte din ce în ce mai mult. Aceasta înseamnă că suma algebrică a abaterilor reziduale $\varepsilon t = y_t - f(t)$ tinde către zero cînd numărul termenilor seriei dinamice tinde către infinit. Este clar că funcţia $f(t)$ va fi cu atît mai reprezentativă pentru evoluţia indicatorului y_t cu cît numărul termenilor seriei dinamice, ce stă la baza estimării acestei funcţii, este mai mare. Necesitatea luării în calculul tendinţei a unui număr suficient de mare de termeni ai unei serii dinamice este uneori uitată în cercetările practice, fapt care transformă operaţia de ajustare într-un simpul exerciţiu de calcul fără nici o semnificaţie reală.

Pentru estimarea valorilor seriei ajustate se pot aplica metode şi tehnici diferite, în funcţie de natura fenomenului cercetat şi de gradul de precizie urmărit.

Prezentăm în continuare cîteva metode dintre cele mai des aplicate în practica statistică.

20.2. Metode de ajustare a seriilor dinamice

20.2.1. Metoda grafică

A ajusta o serie dinamică cu ajutorul metodei grafice înseamnă a trasa vizual tendinţa de dezvoltare a fenomenului supus cercetării şi a obţine astfel o primă imagine asupra formei trendului.

Pentru a ajusta o serie dinamică cu ajutorul metodei grafice este necesar să se reprezinte grafic seria supusă cercetării în cadrul unui sistem de axe rectangulare şi apoi, pe baza distribuţiei punctelor empirice în spaţiul cadranului întîi al axelor de coordonate, se trasează vizual o dreaptă sau o curbă, care să se apropie cît mai mult de toate punctele empirice, fără a trece chiar prin acestea.

Pentru exemplificare, considerăm seria dinamică din tabelul 20.1.

Tabelul 20.1. Producţia de zahăr a României în perioada 1950 - 1977

Anul	Mii tone zahăr	Anul	Mii tone zahăr
1950	87	1964	349
1951	138	1965	402
1952	102	1966	442
1953	115	1967	445
1954	93	1968	384
1955	130	1969	428
1956	151	1970	377
1957	185	1971	484
1958	188	1972	520
1959	242	1973	628
1960	391	1974	516
1961	422	1975	516
1962	342	1976	561
1963	288	1977	713

Sursa: Anuarele statistice ale R.S.România

Se poate observa din fig. 20.1 că tendinţa de creştere a producţiei de zahăr poate fi descrisă satisfăcător, pentru perioada 1950-1977, cu ajutorul unei drepte. Deoarece forma şi poziţia tendinţei s-a obţinut prin trasare vizuală, ea este influenţată, într-o anumită măsură, de experienţa cercetătorului, de subiectivism. De aceea, forma tendinţei, obţinută cu ajutorul metodei grafice, are un caracter orientativ şi se utilizează ca metodă preliminară pentru alegerea tipului de funcţie care să corespundă cel mai bine tendinţei generale de dezvoltare a fenomenelor.

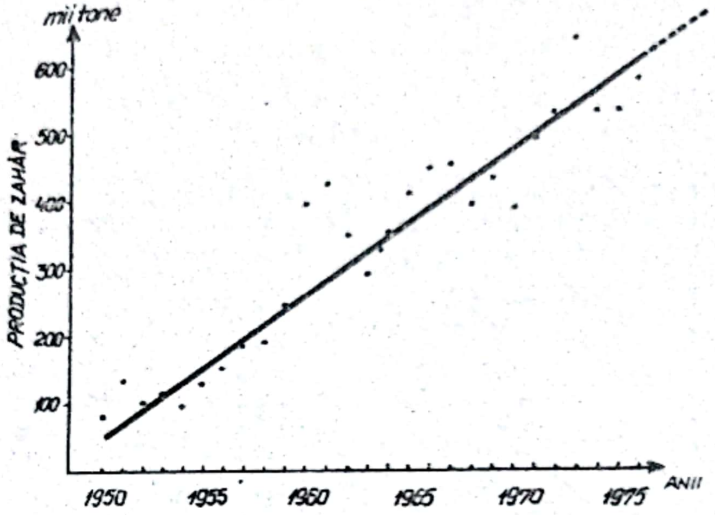


Fig.20.1. PRODUCȚIA DE ZAHĂR A ROMÂNIEI ÎN PERIOADA 1950-1976

În cazul seriilor dinamice, care prezintă oscilaţii sezoniere ciclice, pentru a se înlătura variaţiile sezoniere se poate considera, pentru fiecare ciclu, nivelul mediu al indicatorului y, utilizîndu-se în acest caz metoda mediilor esalonate. Ordonatele egale cu mediile eşalonate vor fi ridicate din punctele de abscisă corespunzătoare centrului ciclurilor de variaţie.

20.2.2. Metoda mediilor mobile

Deseori în practica statistică se întîlnesc serii dinamice a căror tendinţă nu poate fi descrisă cu ajutorul unor funcţii simple, obişnuite, de exemplu, variaţia de forma unei curbe trigonometrice, sinusoidă, cosinusoidă etc., prezentînd anumite oscilaţii al căror sens şi intensitate se repetă ciclic, sezonier. In aceste cazuri se poate utiliza metoda mediilor mobile pentru ajustare.

Dacă ciclurile sînt aproximativ egale ca durată, adică cuprind acelaşi număr de termeni, atunci se recomandă ca numărul termenilor din care se calculează mediile mobile, să fie egal cu numărul de termeni dintr-un ciclu. Egalitatea duratei ciclurilor de variaţie este numai uneori realizată în practică şi, de aceea, în ca-

zul-general, se consideră că numărul termenilor din care se calculează mediile mobile este egal cu numărul de termeni cuprinşi în ciclul de durată medie.

Mediile mobile se calculează ca medii aritmetice pe grupe de termeni, obţinute prin deplasarea succesivă în timp cu câte un termen, prin eliminarea primului termen şi includerea termenului următor ultimului, după formula:

- în momentul t :

$$\bar{y}_t = \frac{y_{t-m} + y_{t-m+1} + \dots + y_t + \dots + y_{t+m-1} + y_{t+m}}{2m+1} \qquad (20.2)$$

- în momentul t+1:

$$\bar{y}_{t+1} = \frac{y_{t-m+1} + y_{t-m+2} + \dots + y_{t+1} + \dots + y_{t+m} + y_{t+m+1}}{2m+1}$$

- în momentul t+2:

$$\bar{y}_{t+2} = \frac{y_{t-m+2} + y_{t-m+3} + \dots + y_{t+2} + \dots + y_{t+m+1} + y_{t+m+2}}{2m+1}$$

.....

Aplicarea în practică a acestei metode de ajustare a seriilor dinamice nu ridică nici un fel de probleme în cazul în care mediile se calculează dintr-un număr impar de termeni: n = 2m+1, deoarece fiecare medie mobilă va cădea în centrul intervalului de timp, chiar pe unul din termenii seriei reale. Astfel, media $\bar{y}_t$ va înlocui termenul y_t, iar $\bar{y}_{t+1}$ va înlocui termenul y_{t+1} şi aşa mai departe (fig.20.2).

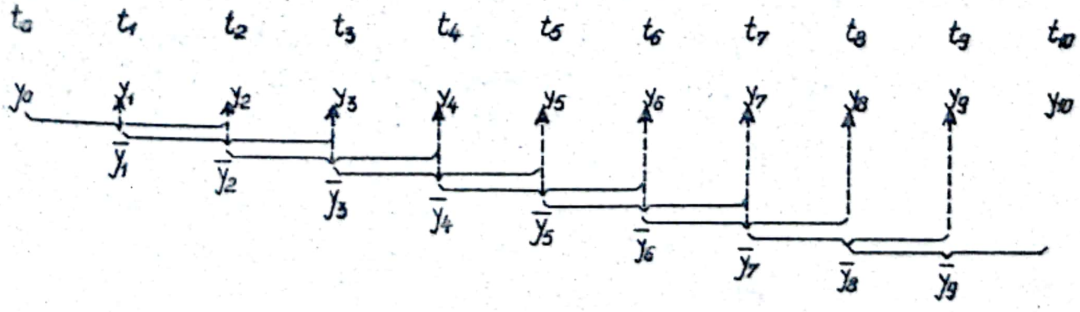


Fig.20.2. AJUSTAREA UNEI SERII DINAMICE CU AJUTORUL METODEI MEDIILOR MOBILE CALCULATE DINTR-UN NUMĂR IMPAR DE TERMENI

In cazul calculării mediilor mobile dintr-un număr par de termeni: n = 2m , mediile vor cădea între doi termeni centrali ai seriei şi va fi nevoie de o nouă operaţie de calcul - numită centrarea mediilor mobile - pentru a determina termenii seriei ajustate, care să cadă chiar în momentele t_1 . Centrarea mediilor mobile înseamnă determinarea mediei aritmetice din cîte două medii mobile consecutive. Schematic, operaţia de ajustare se prezintă în fig. 20.3.

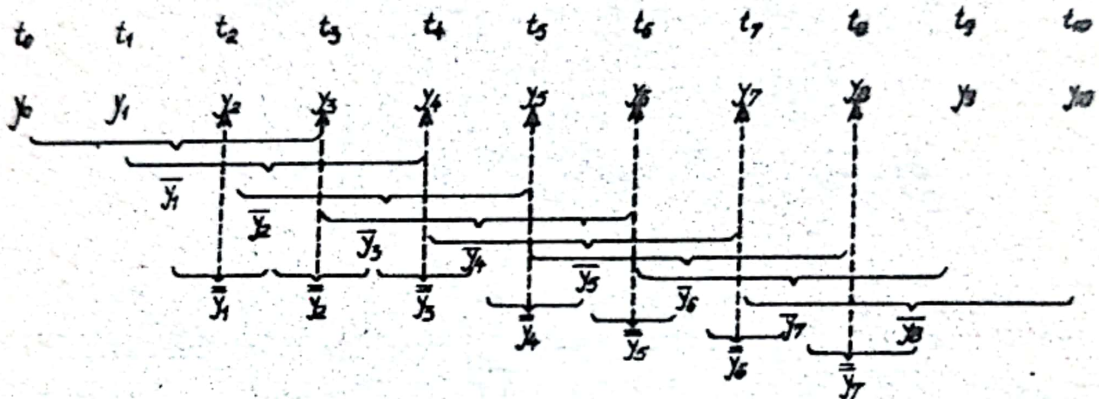


Fig.20.3. AJUSTAREA UNEI SERII DINAMICE CU AJUTORUL METODEI MEDIILOR MOBILE CALCULATE DINTR-UN NUMĂR PAR DE TERMENI

Metoda mediilor mobile, folosită în ajustarea seriilor dinamice, prezintă o serie de dezavantaje inerente.

In primul rînd nu acoperă întreaga perioadă de timp dintr-o serie dinamică reală. Astfel, dacă numărul termenilor k , din care se calculează mediile mobile, este impar, atunci (k-1)/2 valori ale seriei ajustate se pierd la fiecare capăt al seriei reale, iar cînd k este par, atunci se pierd la fiecare capăt al seriei cîte k/2 termeni.

In al doilea rînd, această metodă nu duce la obţinerea unui trend propriu-zis, la o ecuaţie a acestuia, la o "lege" de dezvoltare, ci doar atenuează, într-o anumită măsură, abaterile maxime şi cele minime ale unei variabile.

In al treilea rînd, mediile mobile simple tind să urmeze datele reale spre valorile lor maxime şi minime care au cea mai mare amplitudine şi durată.

In sfîrşit, în unele cazuri, evoluţia obţinută pe baza aplicării metodei mediilor mobile se depărtează foarte mult de evoluţia reală, de exemplu în cazul tendinţei exponenţiale etc.

Toate aceste neajunsuri ale metodei mediilor mobile de ajustare a unei serii dinamice îi îngustează sfera de aplicabilitate practică.

20.2.3. Metoda mediilor înjumătăţite

Această metodă se utilizează în practică pentru determinarea unei tendinţe de formă liniară ce trece prin două puncte. Pentru aceasta, seria dinamică empirică se împarte în două părţi egale. Primul punct este dat de media aritmetică a valorilor lui y, pentru prima jumătate a seriei dinamice şi se află localizat în mijlocul acestei perioade de timp, iar al doilea punct este aritmetică a valorilor lui y din cea de a doua jumătate a seriei şi se află situat în mijlocul perioadei de timp respective.

Parametrii dreptei de tendinţă se determină pe baza formulelor:

$$a = \frac{2S_1}{n} \qquad (20.3)$$

şi

$$b = \frac{4(S_2 - S_1)}{n^2} \qquad (20.4)$$

în care:

S_1 = suma valorilor lui y din prima jumătate a seriei;

S_2 = suma valorilor lui y din a doua jumătate a seriei;

n = numărul termenilor seriei.

In cazul unei serii formate dintr-un număr "n" par de termeni, acesta se poate împărţi în două părţi egale, iar dacă n este impar, atunci termenul din centrul seriei se poate cuprinde în ambele părţi sau se poate neglija.

Pentru exemplificare, considerăm seria dinamică din tabelul 20.1.

Tabelul 20.2. Ajustarea seriei după metoda mediilor înjumătăţite

Anul	Mii tone zahăr		t_i	Seria ajustată $y_{t_i} = a + b.t_i$
1950	87		-6,5	93,36
1951	138		-5,5	110,58
1952	102		-4,5	127,80
1953	115		-3,5	145,02
1954	93		-2,5	162,24
1955	130		-1,5	179,46
1956	151		-0,5	196,68
1957	185	$S_1 = 2874$	0,5	213,90
1958	188		1,5	231,12
1959	242		2,5	248,34
1960	391		3,5	265,56
1961	422		4,5	282,78
1962	342		5,5	300,00
1963	288		6,5	317,22
1964	349		7,5	334,44
1965	402		8,5	351,66
1966	442		9,5	368,88
1967	445		10,5	386,10
1968	384		11,5	403,32
1969	428		12,5	420,54
1970	377		13,5	437,76
1971	484	$S_2 = 6249$	14,5	454,98
1972	520		15,5	472,20
1973	628		16,5	489,42
1974	516		17,5	506,64
1975	576		18,5	523,86
1976	561		19,5	541,08
1977	713		20,5	558,30

Substituind aceste date în formulele parametrilor dreptei de tendinţă (20.3) şi (20.4), rezultă:

$$a = \frac{2S_1}{n} = \frac{2.2874}{28} = 205,29$$

şi

$$b = \frac{4(S_2 - S_1)}{28^2} = 17,22$$

Deci tendinţa liniară de creştere a producţiei de zahăr va fi descrisă de ecuaţia:

$$y_t = 205,29 + 17,22 \cdot t_i$$

Pentru a obține valorile lui t_i considerăm că $y_o = a$ în centrul primei jumătăți a seriei și deci în acel punct $t_i = 0$.

Apoi, în funcție de această origine de calcul, se vor stabili valorile lui t_i, care ulterior vor fi înlocuite succesiv în ecuația de tendință (ultima coloană a tabelului 20.2):

$t = -6,5 \quad y_{-6,5} = 205,29 + 17,22 \cdot (-6,5) = 93,36$

$t = -5,5 \quad y_{-5,5} = 205,29 + 17,22 \cdot (-5,5) = 110,58$

. . . . . .

$t = 20,5 \quad y_{20,5} = 205,29 + 17,22 \cdot (20,5) = 558,30$

20.2.4. Metoda sporului mediu

Metoda sporului mediu se aplică în practica statistică pentru ajustarea seriilor dinamice ale căror termeni variază aproximativ după modelul unei progresii aritmetice. Rația progresiei aritmetice se consideră egală cu sporul mediu de creștere a fenomenului prezentat cu ajutorul seriei dinamice luate în studiu.

Variația unui fenomen în timp se apropie de modelul unei progresii aritmetice dacă sporurile cu bază mobilă sînt apropiate între ele ca mărime. În particular, dacă aceste sporuri sînt chiar egale între ele, atunci dezvoltarea fenomenului cercetat se desfășoară exact după modelul unei expresii aritmetice, situație care se poate întîlni în practică numai în mod cu totul excepțional. Dezvoltarea fenomenelor sociale și economice reale se apropie într-o măsură mai mare sau mai mică de modelul teoretic al unei progresii aritmetice dar, de regulă, nu se va suprapune peste acest model.

Termenii seriei ajustate se obțin pe baza formulei termenului general al unei progresii aritmetice:

$$y_k = y_o + k \bar{\Delta}_y \qquad k = 0,1,2, \ldots , n \qquad (20.5)$$

în care

y_k = un termen oarecare al seriei ajustate;
y_o = nivelul considerat ca bază;
$\bar{\Delta}_y$ = sporul mediu al lui y.

Deoarece ecuaţia (20.5) este similară cu ecuaţia unei drepte, forma tendinţei, obţinută pe baza metodei sporului mediu, va fi întotdeauna liniară.

Se ştie că, în cazul unei progresii aritmetice, y_0 reprezintă nivelul iniţial, cel mai mic ca valoare. Aceasta înseamnă că în cazul ajustării seriilor dinamice, după metoda sporului mediu, se poate considera ca nivel de bază primul termen al seriei reale. In acest caz, prin adăugare în mod succesiv a sporului mediu, se va ajunge în final ca valoarea teoretică a lui y_n să coincidă cu ultimul termen al seriei reale. Deci tendinţa va trece în mod obligatoriu prin punctele extreme ale seriei şi va face abstracţie de toate celelalte intermediare, uneori situîndu-se în totalitate deasupra sau dedesubtul punctelor reale intermediare, diminuînd caracterul de reprezentativitate al tendinţei. De aceea, acest procedeu se recomandă a se aplica în practică numai în cazurile în care, fie că nu se cunosc decît valorile extreme ale unei serii dinamice, fie că se cunosc şi termenii intermediari, dar nivelurile extreme reprezintă niveluri de dezvoltare normală a fenomenului.

In toate celelalte cazuri este necesar să se identifice, prin analiza seriei, care este nivelul ce se apropie în cea mai mare măsură de tendinţa normală de dezvoltare şi apoi să-l considerăm ca nivel de bază y_0. Un procedeu simplu de alegere a nivelului de bază constă în reprezentarea grafică a seriei urmată de trasarea vizuală a tendinţei. Ordonata punctului care se va apropia în cea mai mare măsură de tendinţa astfel trasată se va considera nivel de bază y_0.

Considerăm, pentru exemplificare, seria dinamică din tabelul 20.3. Observăm, din coloana trei a tabelului, că sporurile absolute cu bază mobilă sînt apropiate între ele ca mărime, ceea ce constituie un argument în favoarea ipotezei că evoluţia producţiei de ţiţei se apropie de modelul unei progresii aritmetice avînd raţia egală cu :

$$\bar{\Delta}_y = \frac{y_n - y_0}{n} = \frac{14,7 - 11,5}{16} = \frac{3,2}{16} = 0,2 \text{ milioane tone}$$

In cazul acestei serii, nici un termen nu se abate în mare

Tabelul 20.3. Producția de țiței extras în România în perioada 1960 - 1976 și prognoza pînă în anul 1985

Anii	Țiței extras (milioane tone) (y_i)	$\Delta y_{i/i-1}$	k	Seria ajustată $y_k = 11,5 + 0,2k$	$\lvert y_i - y_k \rvert$	k'	Seria ajustată $y_{k'} = 13,8 + 0,2k'$	$\lvert y_i - y_{k'} \rvert$
1	2	3	4	5	6	7	8	9
1960	11,5	-	0	11,5	0	-11	11,6	0,1
1961	11,6	0,1	1	11,7	0,1	-10	11,8	0,2
1962	11,9	0,3	2	11,9	0	-9	12,0	0,1
1963	12,2	0,3	3	12,1	0,1	-8	12,2	0
1964	12,4	0,2	4	12,3	0,1	-7	12,4	0
1965	12,6	0,2	5	12,5	0,1	-6	12,6	0
1966	12,8	0,2	6	12,7	0,1	-5	12,8	0
1967	13,2	0,4	7	12,9	0,3	-4	13,0	0,2
1968	13,3	0,1	8	13,1	0,2	-3	13,2	0,1
1969	13,2	0,1	9	13,3	0,1	-2	13,4	0,2
1970	13,4	0,2	10	13,5	0,1	-1	13,6	0,2
1971	13,8	0,4	11	13,7	0,1	0	13,8	0
1972	14,1	0,3	12	13,9	0,2	1	14,0	0,1
1973	14,3	0,2	13	14,1	0,2	2	14,2	0,1
1974	14,5	0,2	14	14,3	0,2	3	14,4	0,1
1975	14,6	0,1	15	14,5	0,1	4	14,6	0
1976	14,7	0,1	16	14,7	0	5	14,8	0,1
Total	224,1	3,2	-	222,7	2,0	-	224,4	1,5
1977	-	-	17	14,9	-	6	15,0	-
1978	-	-	18	15,1	-	7	15,2	-
1979	-	-	19	15,3	-	8	15,4	-
1980	-	-	20	15,5	-	9	15,6	-
1981	-	-	21	15,7	-	10	15,8	-
1982	-	-	22	15,9	-	11	16,0	-
1983	-	-	23	16,1	-	12	16,2	-
1984	-	-	24	16,3	-	13	16,4	-
1985	-	-	25	16,5	-	14	16,6	-

Scanned with OKEN Scanner

măsură de la creşterea normală a producţiei de ţiţei extras şi deci, putem considera ca origine de calcul, primul termen al seriei $y_0 = 11,5$.

Substituind pe y_0 şi $\bar{\Delta}_y$ în formula (20.5) rezultă termenii seriei ajustate (col. 5, tabelul 20.3):

$$y_k = 11,5 + k \cdot 0,2$$

pentru $k = 0$ $\quad y_0 = 11,5$

$k = 1$ $\quad y_1 = 11,5 + 1 \cdot 0,2 = 11,7$

$k = 2$ $\quad y_2 = 11,5 + 2 \cdot 0,2 = 11,9$

. . . . . .

$k = 16$ $\quad y_{16} = 11,5 + 16 \cdot 0,2 = 14,7$

Reprezentînd şi tendinţa pe acelaşi grafic cu seria empirică, se constată că abaterile pozitive ale valorilor reale ale lui y_i sînt mai mari decît abaterile negative, ceea ce înseamnă că seria teoretică nu este cea mai potrivită pentru seria reală. Această nepotrivire se datorează faptului că originea de calcul y_0 s-a ales egală cu primul termen al seriei, exagerîndu-se astfel importanţa acestuia.

Ajustînd vizual seria şi alegînd pentru y_0 acea valoare a varibilei, care se apropie în cea mai mare măsură de tendinţa trasată vizual, adică $y_0 = 13,8$ corespunzător anului 1971, va rezulta o tendinţă mult mai apropiată de seria reală, coloana 8 a tabelului 20.3, care este mai reprezentativă pentru evoluţia producţiei de ţiţei.

Gradul de reprezentativitate a tendinţei unei serii dinamice se poate aprecia cu ajutorul expresiei $\sum |y_i - y_t|$ din coloanele 6 şi 9. De asemenea, observăm că, în cel de al doilea caz, suma termenilor seriei ajustate este mai apropiată de suma termenilor seriei empirice decît în prima ajustare.

20.2.5. Metoda ritmului mediu

Metoda ritmului mediu se poate utiliza în practică pentru ajustarea seriilor dinamice ale căror termeni variază aproximativ

după modelul unei progresii geometrice, iar expresia ei grafică ia forma unei curbe exponenţiale. In general, se apreciază că o serie dinamică prezintă variaţia unui fenomen în progresie geometrică dacă ritmurile variaţiei cu bază mobilă se apropie ca mărime între ele.

Termenii seriei ajustate se determină pe baza formulei termenului general al unei progresii geometrice:

$$y_k = y_o \cdot \bar{R}_y^k \quad ; \quad k = 0,1,2,\ldots,n \qquad (20.6)$$

în care

y_k = un termen oarecare al progresiei geometrice;
y_o = nivelul lui y considerat ca origine de calcul;
$\bar{R}_y$ = ritmul mediu de variaţie al seriei dinamice reale;
k = numărul de termeni care separă pe y_k de y_o.

Ca şi în cazul metodei sporului mediu, în aplicarea acestei metode, originea de calcul y_o se poate considera ca fiind egală cu nivelul iniţial al lui y din cadrul seriei, sau mai bine, se alege pe baza ajustării vizuale . ca fiind egal cu acel nivel al seriei care se apropie în cea mai mare măsură de tendinţa trasată vizual.

Pentru exemplificare, considerăm evoluţia producţiei de oţel a României în perioada 1950-1976 (vezi tabelul 20.4). Reprezentînd grafic această serie (fig. 20.4) se observă că producţia de oţel evoluează aproximativ după o curbă exponenţială şi că unor intervale egale de pe axa timpului le corespund creşteri din ce în ce mai mari ale producţiei de oţel.

Ritmul mediu anual al creşterii producţiei de oţel în perioada 1950-1976 va fi:

$$\bar{R} = \sqrt[26]{\frac{10732}{555}} = 1,121 \text{ ori} \quad \text{sau} \quad \bar{r} = 12,1\%$$

Alegînd drept origine de calcul primul termen al seriei $y_o = 555$, corespunzător anului 1950, pe col. 3 a tabelului 20.4, s-au determinat valorile succesive ale lui k.

Substituind în formula (20.6) ritmul mediu şi k, rezultă termenii seriei ajustate din col. 4 a tabelului 20.4.

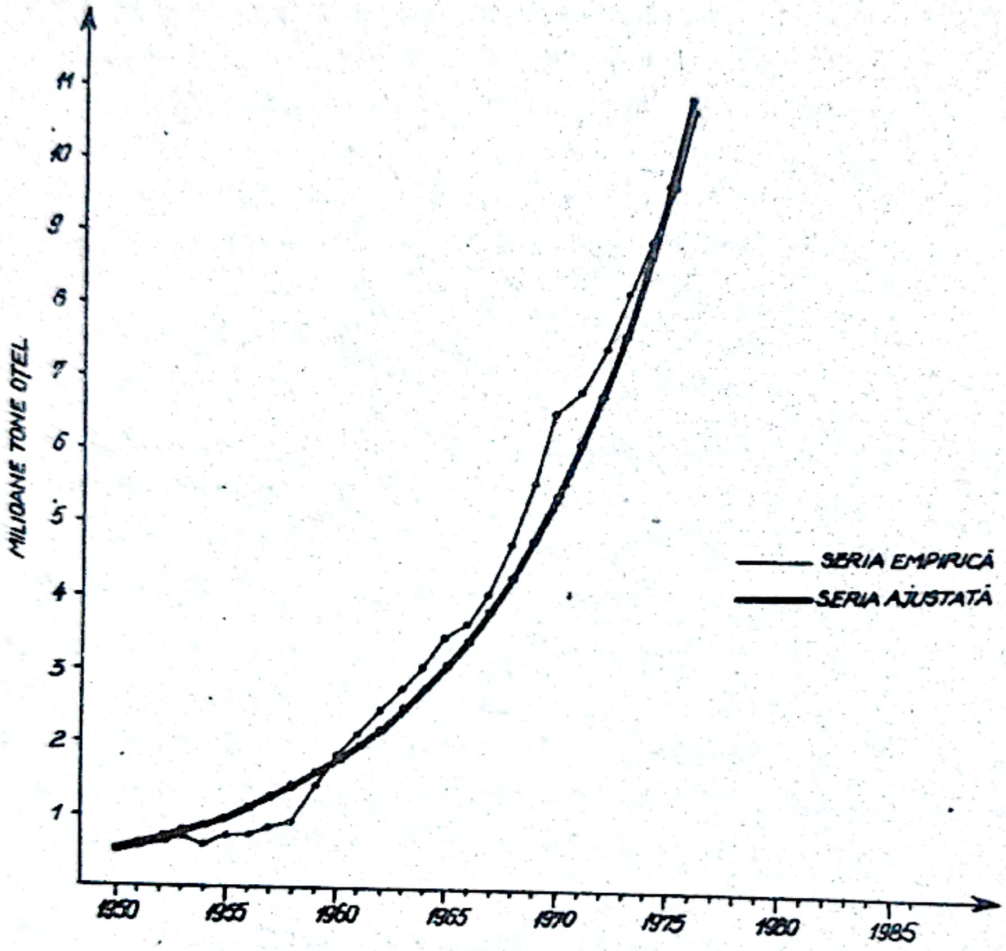


Fig.20.4. PRODUCȚIA DE OȚEL A ROMÂNIEI ÎN PERIOADA 1950 - 1976

Scanned with OKEN Scanner

Tabelul 20.4. Ajustarea seriei dinamice cu ajutorul metodei ritmului mediu

Anii	Mii tone (y_i)	k	Seria ajustată $y_k = y_0 \cdot \bar{R}^k$	$\lvert y_i - y_k \rvert$
1	2	3	4	5
1950	555	0	555	0
1951	644	1	622	22
1952	698	2	697	1
1953	717	3	782	65
1954	628	4	876	248
1955	766	5	982	216
1956	779	6	1101	322
1957	864	7	1235	371
1958	934	8	1384	450
1959	1420	9	1551	131
1960	1806	10	1739	67
1961	2126	11	1950	176
1962	2451	12	2186	265
1963	2704	13	2450	254
1964	3039	14	2746	293
1965	3426	15	3078	348
1966	3670	16	3451	219
1967	4088	17	3869	219
1968	4751	18	4337	414
1969	5540	19	4862	678
1970	6517	20	5450	1067
1971	6803	21	6110	693
1972	7401	22	6849	552
1973	8161	23	7678	483
1974	8848	24	8607	241
1975	9549	25	9648	99
1976	10733	26	10815	82
Total	99618	-	95610	7976

Sursa: Anuarele statistice ale R.S.R.

20.2.6. Metoda celor mai mici pătrate

Principiul de bază al celor mai mici pătrate a fost prezentat în capitolul referitor la regresie şi corelaţie. Problema fundamentală care trebuie rezolvată în ajustarea unei serii dinamice, folosind metoda celor mai mici pătrate, este alegerea tipului celui mai potrivit de funcţie $f(t)$, pe baza analizei fenomenului luat în calcul şi în concordanţă cu natura lui.

Funcţiile mai des utilizate în practica statistică pentru

determinarea tendinţei generale de dezvoltare a fenomenelor sînt:

1) funcţii polinomiale:

a) polinomul de gradul 1:

$$y = a + bt \tag{20.7}$$

b) polinomul de gradul 2:

$$y = a + bt + ct^2 \tag{20.8}$$

c) polinomul de gradul 3:

$$y = a + bt + ct^2 + dt^3 \tag{20.9}$$

.....

d) polinomul de gradul n :

$$y = a + bt + ct^2 + dt^3 + et^4 + \ldots\ldots + \omega t^n \tag{20.10}$$

2) hiperbola echilateră:

$$y = a + \frac{b}{x} \tag{20.11}$$

3) funcţia exponenţială:

$$y = ab^t \tag{20.12}$$

4) funcţia exponenţială modificată:

$$y = k + ab^t \tag{20.13}$$

5) curba lui Gompertz:

$$y = ab^{c^t} \tag{20.14}$$

6) curba logistică:

$$y = \frac{b}{1 + e^{-\beta t - c}} \tag{20.15}$$

20.2.6.1. Tendinţa liniară

Tipul de funcţie care descrie tendinţa liniară de dezvoltare a fenomenelor este dat de ecuaţia generală a unei drepte:

$$y = a + bt$$

Determinarea parametrilor acestei funcţii, cu ajutorul metodei celor mai mici pătrate, se face prin rezolvarea sistemului:

Scanned with OKEN Scanner

$$na + b\sum t_i = \sum y_i$$
$$a\sum t_i \; b\sum t_i^2 = \sum t_i y_i \qquad (20.16)$$

Deoarece t_i , în cazul unei serii dinamice, variază în progresie aritmetică cu raţia egală cu 1 , se poate face o mişcare de translaţie a dreptei faţă de originea axelor de coordonate în aşa fel, încît centrul intervalului de timp, luat în calcul, să coincidă cu originea axelor de coordonate (fig. 20.5).

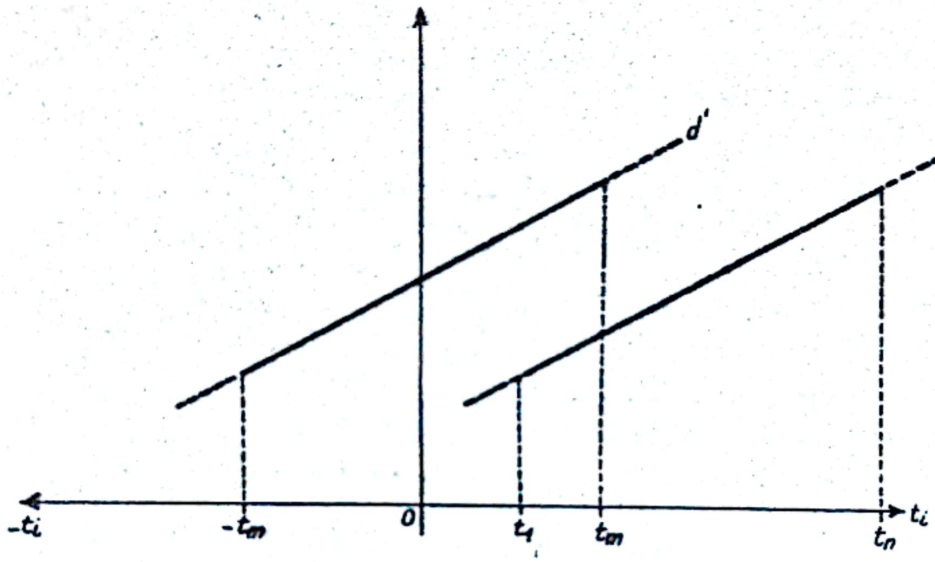


Fig. 20.5. MIŞCAREA DE TRANSLAŢIE A DREPTEI FAŢĂ DE ORIGINE

Abscisele diferitelor puncte ale dreptei d' se exprimă în abateri faţă de anul din centrul seriei, considerat drept origine de calcul. Intrucît abscisa punctului din centrul seriei ("anul fals") este egală cu nivelul mediu al variabilei independente, rezultă, pe baza proprietăţilor mediei aritmetice, că suma abaterilor pozitive este egală în valoare absolută cu suma abaterilor negative ,adică :

$$\sum t_i = 0 \qquad (20.17)$$

Substituind relaţia (20.17) în sistemul (20.16),rezultă:

$$\begin{aligned} na &= \sum y_i \\ b\sum t_i^2 &= \sum t_i y_i \end{aligned} \qquad (20.18)$$

Rezolvînd sistemul (20.18) se obţin formulele de calcul simplificat al celor doi parametri ai ecuaţiei liniare:

$$a = \frac{\sum y_i}{n} \qquad (20.19)$$

$$b = \frac{\sum t_i y_i}{\sum t_i^2} \qquad (20.20)$$

Aplicarea în practică a formulelor (20.19) şi (20.20) se justifică numaiîn cazul îndeplinirii relaţiei (20.17). Alegerea originii de calcul se face în mod diferit după cum numărul termenilor seriei ce urmează a fi ajustată este par sau impar.

In cazul unei serii formate dintr-un număr impar de termeni, originea de calcul va coincide chiar cu termenul din centrul seriei, iar abaterile faţă de origine se vor exprima în ani.

Considerăm, pentru exemplificare, datele din tabelul 20.1.

Analizînd seria, observăm că producţia de zahăr, în perioada 1951-1977 a înregistrat o evoluţie neuniformă, crescătoare ca tendinţă. Din fig. 20.1 se observă că forma tendinţei se poate aproxima cu ajutorul unei drepte. Parametrii dreptei de tendinţă îi vom determina cu ajutorul metodei celor mai mici pătrate, utilizînd formulele de calcul simplificat.

Seria dinamică din tabelul 20.5 cuprinde un număr impar de termeni şi anume $n = 27$. In centrul ei se va găsi termenul al $m+1$-lea, iar deasupra şi dedesubtul lui se vor găsi cîte m termeni, unde $m = (n-1)/2 = (27-1)/2 = 13$. Deci în centrul seriei se află cel de-al 14-lea termen, corespunzător anului 1964. Aceasta înseamnă că anul 1964 se va considera ca an de bază, iar pentru determinarea valorilor succesive ale lui t_i (de pe coloana 3 a tabelului 20.5) se va scădea din fiecare an, anul de bază.

Substituind aceste date în formulele (20.19) şi (20.20), rezultă:

$$a = \frac{\sum y_i}{n} = \frac{9552}{27} = 353,78 \text{ mii tone}$$

$$b = \frac{\sum y_i t_i}{\sum t_i^2} = \frac{33570}{1638} = 20,49 \text{ mii tone}$$

Valoarea $a = 353,78$ mii tone înseamnă că, în perioada 1951-1977, producţia medie anuală de zahăr a României a fost de 353,78 mii tone, iar $b = 20,49$ mii tone semnifică faptul că producţia de zahăr a crescut în medie, în fiecare an al perioadei 1951-1977, cu 20,49 mii tone.

Deci, putem afirma că producţia de zahăr a urmat, în perioada analizată, modelul liniar descris de ecuaţia:

$$y_i = 353,78 + 20,49 \cdot t_i$$

Această dreaptă intersectează axa ordonatelor în punctul cu $y = 353,78$ şi are un coeficient unghiular egal cu 20,49. Menţionăm faptul că ea nu s-a determinat pe baza valorilor reale ale lui t_i, ci cu ajutorul abaterilor acestora faţă de anul din mijloc. Dreapta reală va fi cea care îşi va muta din nou centrul, din origine în anul 1964.

Substituind succesiv valorile lui t_i în ecuaţia dreptei se vor obţine valorile ajustate ale producţiei de zahăr (vezi col. 6 a tabelului 20.5).

Valorile reale şi cele teoretice ale producţiei de zahăr se reprezintă grafic în fig. 20.6.

Scanned with OKEN Scanner

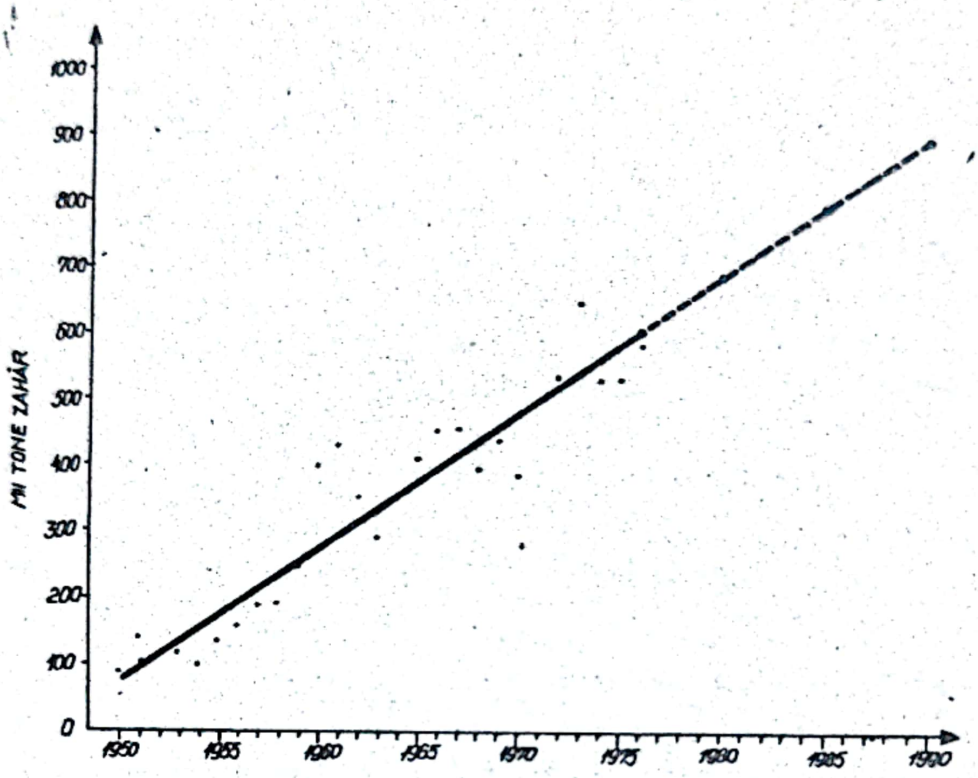


Fig. 20.6. PRODUCȚIA DE ZAHĂR A ROMÂNIEI ÎN PERIOADA 1950-1976 ȘI PROGNOZA PÎNĂ ÎN ANUL 1990

Pentru a determina volumul viitor al producției de zahăr, în ipoteza că și în viitor aceasta va evolua pe baza aceluiași model de creștere, este suficient să substituim valorile lui t_i pentru anii de perspectivă în ecuația de tendință, astfel:

în 1980, $t_i = 16$; $y_{1980} = 681,62$ mii tone
în 1985, $t_i = 21$; $y_{1985} = 784,07$ mii tone
în 1990, $t_i = 26$; $y_{1990} = 886,52$ mii tone , etc.

Din fig. 20.6 se observă că aceleași valori s-ar fi putut obține mai ușor pe cale grafică.

In cazul seriilor cu un număr par de termeni, centrul intervalului de timp luat în calcul va coincide cu mijlocul intervalului de timp dintre cei doi termeni centrali ai seriei. Dacă

Tabelul 20.5. Calculul tendinței liniare cu ajutorul metodei celor mai mici pătrate a seriei reprezentînd producția de zahăr a României în perioada 1951-1977

Anii	Mii tone zahăr (y_i)	t_i	t_i^2	$t_i y_i$	$y_t = a+bt$	$y_i - y_{t_i}$
1	2	3	4	5	6	7
1951	138	-13	169	-1754	87,41	50,59
1952	102	-12	144	-1224	107,90	-5,90
1953	115	-11	121	-1265	128,39	-13,39
1954	93	-10	100	-930	148,88	-55,88
1955	130	-9	81	-1170	169,37	-39,37
1956	151	-8	64	-1208	189,86	-38,86
1957	185	-7	49	-1295	210,35	-25,35
1958	188	-6	36	-1128	230,84	-42,84
1959	242	-5	25	-1210	251,33	-9,33
1960	391	-4	16	-1564	271,82	119,18
1961	422	-3	9	-1266	292,31	129,69
1962	342	-2	4	-684	312,80	29,20
1963	288	-1	1	-288	333,29	-45,29
1964	349	0	0	-15026	353,78	-4,78
1965	402	1	1	402	374,27	27,73
1966	442	2	4	884	394,76	47,24
1967	445	3	9	1335	415,25	29,75
1968	384	4	16	1536	435,74	-51,74
1969	428	5	25	2140	456,23	-28,23
1970	377	6	36	2262	476,72	-99,72
1971	484	7	49	3388	497,21	-13,21
1972	520	8	64	4160	517,70	2,30
1973	628	9	81	5652	538,19	89,81
1974	516	10	100	5160	558,68	-42,68
1975	516	11	121	5676	579,17	-63,17
1976	561	12	144	6732	599,66	-38,66
1977	713	13	169	9269	620,15	92,85
Total	9552	0	1638	33570	9552,06	-0,06

Scanned with OKEN Scanner

vom scădea din fiecare an originea de calcul, atunci abaterile faţă de origine, din diferiţi ani se vor exprima în jumătăţi de an (în cazul general, în care indicatorii y sînt anuali).

Pentru exemplificare, considerăm din nou seria anterioară, la care adăugăm producţia din anul 1950 şi vom avea deci $n = 28$ termeni. Originea de calcul, în acest caz, va fi $t_0 = 1963,5$.

Înlocuind în formulele (20.19) şi (20.20) valorile intermediare obţinute pe baza tabelului 20.6, se obţin valorile parametrilor a şi b :

$$a = \frac{\sum y_i}{n} = \frac{9639}{28} = 344,25 \text{ mii tone}$$

şi

$$b = \frac{\sum t_i y_i}{\sum t_i^2} = \frac{74388}{7308} = 10,18 \text{ mii tone}$$

Deci, producţia de zahăr a României a crescut în perioada 1950-1970 după modelul liniar:

$$y_i = 344,25 + 10,18 \cdot t_i$$

Parametrii acestei ecuaţii au următoarele semnificaţii: $a = 344,25$ înseamnă că în perioada 1950-1977 producţia medie anuală de zahăr a României a fost egală cu 344,25 mii tone; $b = 10,18$ arată că, în fiecare jumătate de an, în medie, producţia de zahăr a crescut în aceeaşi perioadă cu 10,18 mii tone sau anual cu $2b = 20,36$ mii tone.

Comparînd valorile tendinţei liniare, obţinute în tabelele 20.5 şi 20.6 observăm că ele sînt foarte apropiate, fapt logic dacă ne gîndim că în primul tabel avem aceeaşi serie ca în al doilea cu excepţia primului termen.

Tabelul 20.6. Calculul tendinței liniare cu ajutorul metodei celor mai mici pătrate, a seriei reprezentînd producția de zahăr a României în perioada 1950 - 1977

Anii	Mii tone zahăr (y_i)	t_i	t_i^2	$t_i y_i$	$y_t = a + bt$	$y_i - y_t$
1	2	3	4	5	6	7
1950	87	-27	729	-2349	69,39	17,61
1951	138	-25	625	-3450	89,75	48,25
1952	102	-23	529	-2346	110,11	-8,11
1953	115	-21	441	-2415	130,47	-15,47
1954	93	-19	361	-1767	150,83	-57,83
1955	130	-17	289	-2210	171,19	-41,19
1956	151	-15	225	-2265	191,55	-40,55
1957	185	-13	169	-2360	211,91	-26,91
1958	188	-11	121	-2068	232,27	-44,27
1959	242	-9	81	-2178	252,63	-10,63
1960	391	-7	49	-2737	272,99	118,01
1961	422	-5	25	-2110	293,35	128,65
1962	342	-3	9	-1026	313,71	28,29
1963	288	-1	1	-288	334,07	-46,07
1964	349	1	1	349	354,43	-5,43
1965	402	3	9	1206	374,79	27,21
1966	442	5	25	2210	395,15	46,85
1967	445	7	49	3115	415,51	29,49
1968	384	9	81	3456	435,87	-51,87
1969	423	11	121	4708	456,23	-28,23
1970	377	13	169	4901	476,59	-99,59
1971	484	15	225	7260	496,95	-12,95
1972	520	17	289	8840	517,31	2,69
1973	628	19	361	11932	537,67	90,33
1974	516	21	441	10836	558,03	-42,03
1975	516	23	529	11868	578,39	-62,39
1976	561	25	625	14025	598,75	-37,75
1977	713	27	729	19251	619,11	93,89
Total	9639	0	7308	74388	9639,00	0,00

Comparînd valorile tendinței liniare, obținute în tabelele 20.5 și 20.6 observăm că ele sînt foarte apropiate, fapt logic dacă ne gîndim că în primul tabel avem aceeași serie ca în al doilea, cu excepția primului termen.

20.2.6.2. Tendința după o parabolă de gradul doi

Aplicarea metodei celor mai mici pătrate pentru estimarea parametrilor de gradul doi duce la obținerea următorului sistem

de ecuaţii normale:

$$\begin{aligned} na + b\sum t_i + c\sum t_i^2 &= \sum y_i \\ a\sum t_i + b\sum t_i^2 + c\sum t_i^3 &= \sum t_i y_i \\ a\sum t_i^2 + b\sum t_i^3 + c\sum t_i^4 &= \sum t_i^2 y_i \end{aligned} \qquad (20.21)$$

Ca şi în cazul tendinţei liniare, calculele parametrilor tendinţei parabolice se pot simplifica făcînd substituţia $\sum t_i = 0$; $\sum t_i^3 = 0$ în sistemul (20.21):

$$\begin{aligned} na + c\sum t_i^2 &= \sum y_i \\ b\sum t_i^2 &= \sum t_i y_i \\ a\sum t_i^2 + c\sum t_i^4 &= \sum t_i^2 y_i \end{aligned} \qquad (20.22)$$

Din ecuaţia a doua a sistemului (20.22) rezultă:

$$b = \frac{\sum t_i y_i}{\sum t_i^2} \qquad (20.23)$$

Rezolvînd sistemul format din prima şi a treia ecuaţie din (20.22), rezultă formulele de calcul ale parametrilor a şi c :

$$a = \frac{\sum t_i^4 \sum y_i - \sum t_i^2 \sum t_i^2 y_i}{n\sum t_i^4 - (\sum t_i^2)^2} \qquad (20.24)$$

şi

$$c = \frac{n\sum t_i^2 y_i - \sum t_i^2 \sum y_i}{n\sum t_i^4 - (\sum t_i^2)^2} \qquad (20.25)$$

Alegerea originii de calcul în aşa fel încît $\sum t_i = 0$ şi $\sum t_i^3 = 0$ se face în acelaşi fel ca şi în cazul tendinţei liniare.

Pentru exemplificare, să considerăm evoluţia populaţiei municipiului Iaşi în perioada 1950-1976 (tabelul 20.7).

Inlocuind valorile din tabel în formulele (20.23),(20.24) şi (20.25), rezultă valorile parametrilor a,b şi c, astfel:

$$a = \frac{\sum t_i^4 \sum y_i - \sum t_i^2 \sum t_i^2 y_i}{n\sum t_i^4 - (\sum t_i^2)^2} = \frac{178542.3925 - 1638.252207}{27.178542 - 1638^2} = 134,5732 \text{ mii loc.}$$

$$b = \frac{\sum t_i y_i}{\sum t_i^2} = \frac{7587}{1638} = 4,6319 \text{ mii loc.}$$

$$c = \frac{n\sum t_i^2 y_i - \sum t_i^2 \sum y_i}{n\sum t_i^4 - (\sum t_i^2)^2} = \frac{27.252207 - 1638.3925}{27.178542 - 1638^2} = 0,1780 \text{ mii loc.}$$

Aceasta înseamnă că evoluţia populaţiei municipiului Iaşi, în perioada 1950-1976, s-a produs după modelul unei parabole de gradul doi:

$$y_i = 134,57 + 4,63\ t_i + 0,18\ t_i^2$$

Dacă substituim succesiv, în această formulă, valorile lui t_i din col.3 a tabelului 20.7, se obţin valorile modelului parabolic din col. 8 a aceluiaşi tabel.

Pentru a determina numărul populaţiei viitoare a municipiului Iaşi se substituie, în ecuaţia de tendinţă, valorile pe care le va lua t_i în anii viitori, astfel:

- în anul 1980: $t_{1980} = 17$; $y_{1980} = 265,3$ mii loc.
- în anul 1985: $t_{1985} = 22$; $y_{1985} = 323,55$ mii loc.
- în anul 1990: $t_{1990} = 27$; $y_{1990} = 390,80$ mii loc.
- în anul 2000: $t_{2000} = 37$; $y_{2000} = 552,3$ mii loc.

Pentru a se obţine o simplificare mai mare a calculelor la însumarea puterilor pare ale lui t_i, în sistemul (20.22) se poate face substituţia:

$$\sum_{-m}^{m} t_i^{2k} = 2 \sum_{0}^{m} t_i^{2k} \qquad (20.26)$$

unde $n = 2m + 1$,

Sistemul (20.22) devine:

$$na + 2c\sum_{0}^{m} t_i^2 = \sum_{-m}^{m} y_i$$

$$2b \sum_{0}^{m} t_i^2 = \sum_{-m}^{m} t_i y_i \qquad (20.27)$$

$$2a \sum_{0}^{m} t_i^2 + 2e \sum_{0}^{m} t_i^4 = \sum_{-m}^{m} t_i^2 y_i$$

Rezolvînd sistemul (20.27) rezultă formulele parametrilor parabolei de gradul doi, cu ajutorul cărora se obţine cea mai mare simplificare a calculelor:

$$a = \frac{2 \sum_{0}^{m} t_i^4 \sum_{-m}^{m} y_i - 2 \sum_{0}^{m} t_i^2 \sum_{-m}^{m} t_i^2 y_i}{2n\sum_{0}^{m} t_i^4 - (2 \sum_{0}^{m} t_i^2)^2} \qquad (20.28)$$

$$b = \frac{\sum_{-m}^{m} t_i y_i}{2\sum_{0}^{m} t_i^2} \qquad (20.29)$$

$$c = \frac{n\sum_{-m}^{m} t_i^2 y_i - 2\sum_{0}^{m} t_i^2 \sum_{-m}^{m} y_i}{2n\sum_{0}^{m} t_i^4 - (2 \sum_{0}^{m} t_i^2)^2} \qquad (20.30)$$

Tabelul 20.7. Evoluția populației municipiului Iași în perioada 1950 - 1976 și prognoza pînă în anul 1990

Anii	Mii locuitori (y_i)	t_i	t_i^2	t_i^4	$t_i y_i$	$t_i^2 y_i$	$y=a+bt+ct^2$
							8
1950	102	-13	169	28561	-1326	17238	104
1951	104	-12	144	20736	-1248	14976	105
1952	105	-11	121	14641	-1155	12705	105
1953	107	-10	100	10000	-1070	10700	106
1954	108	-9	81	6561	-972	8748	107
1955	111	-8	64	4096	-888	7104	109
1956	113	-7	49	2401	-791	5537	111
1957	118	-6	36	1296	-708	4248	113
1958	121	-5	25	625	-605	3025	116
1959	121	-4	16	256	-484	1936	119
1960	122	-3	9	81	-366	1098	122
1961	124	-2	4	16	-248	496	126
1962	125	-1	1	1	-125	125	130
1963	126	0	0	0	0	0	135
1964	126	1	1	1	126	126	139
1965	124	2	4	16	248	496	145
1966	161	3	9	81	483	1449	150
1967	164	4	16	256	656	2624	156
1968	170	5	25	625	850	4250	162
1969	176	6	36	1296	1056	6336	169
1970	182	7	49	2401	1274	8918	176
1971	186	8	64	4096	1488	11904	183
1972	191	9	81	6561	1719	15471	191
1973	198	10	100	10000	1980	19800	199
1974	206	11	121	14641	2266	24926	207
1975	215	12	144	20736	2580	30960	216
1976	219	13	169	28561	2847	37011	225
Total	3925	0	1638	178542	7587	252207	3925
1977	-	14	196	-	-	-	234
1978	-	15	225	-	-	-	244
1979	-	16	256	-	-	-	254
1980	-	17	289	-	-	-	265
1981	-	18	324	-	-	-	276
1982	-	19	361	-	-	-	287
1983	-	20	400	-	-	-	298
1984	-	21	441	-	-	-	310
1985	-	22	484	-	-	-	324
.....	.....	.....	.....	.....	.....	.....	.....
1990	-	27	729	-	-	-	391
.....	.....	.....	.....	.....	.....	.....	.....
2000	-	37	1369	-	-	-	552

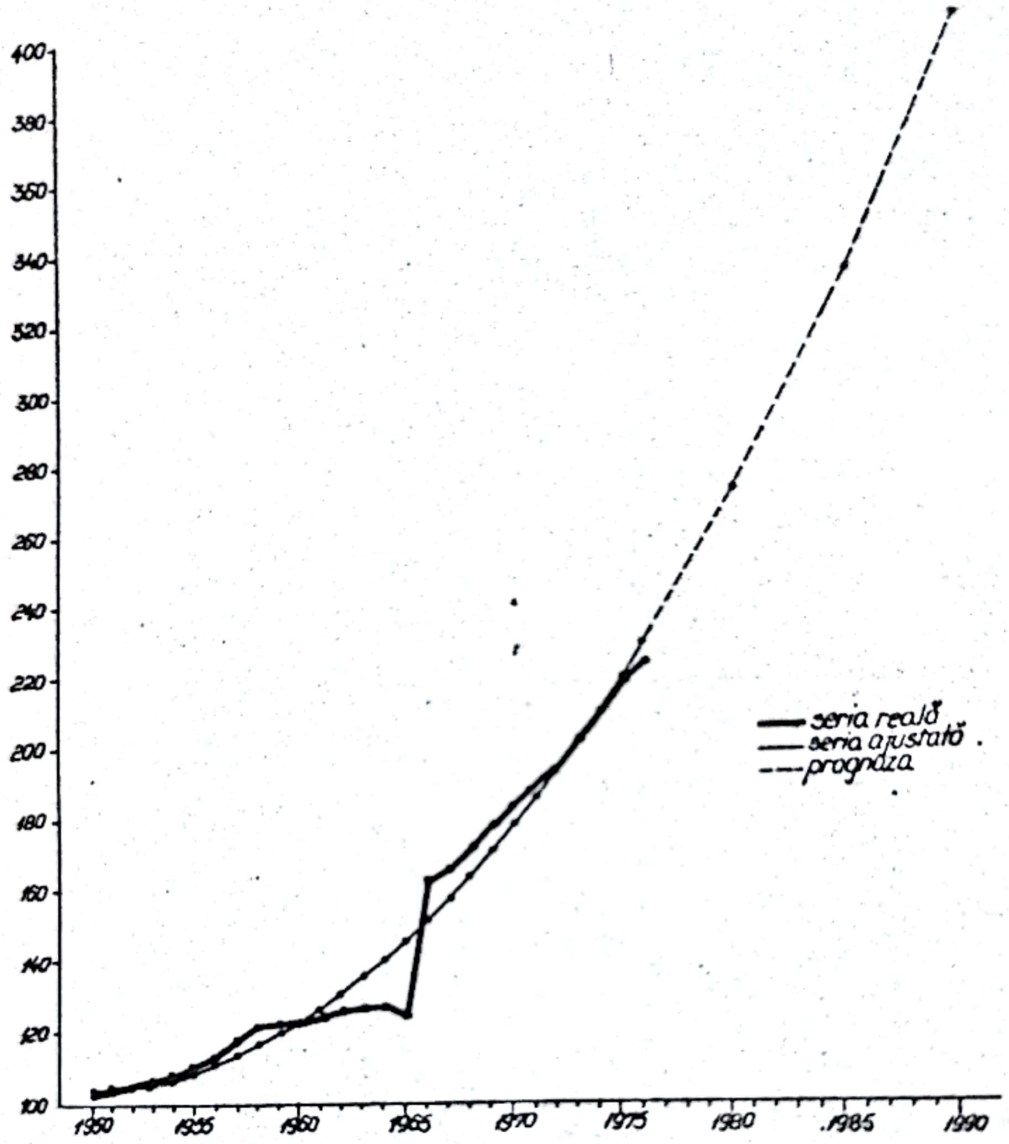


Fig 20.7. EVOLUȚIA POPULAȚIEI MUNICIPIULUI IAȘI ÎN PERIOADA 1950-1976 ȘI PROGNOZA PÎNĂ ÎN ANUL 1990

20.3.6.3. Tendința după o parabolă de gradul 3

Determinarea parametrilor ecuației de tendință, cu ajutorul metodei celor mai mici pătrate în cazul parabolei de gradul 3, presupune rezolvarea sistemului de ecuații normale:

$$\begin{aligned} na + b\sum t_i + c\sum t_i^2 + d\sum t_i^3 &= \sum y_i \\ a\sum t_i + b\sum t_i^2 + c\sum t_i^3 + d\sum t_i^4 &= \sum t_i y_i \\ a\sum t_i^2 + b\sum t_i^3 + c\sum t_i^4 + d\sum t_i^5 &= \sum t_i^2 y_i \\ a\sum t_i^3 + b\sum t_i^4 + c\sum t_i^5 + d\sum t_i^6 &= \sum t_i^3 y_i \end{aligned} \qquad (20.31)$$

Deoarece sumele puterilor impare ale lui t_i sînt nule, sistemul (20.31) devine:

$$\begin{aligned} na + c\sum t_i^2 &= \sum y_i \\ b\sum t_i^2 + d\sum t_i^4 &= \sum t_i y_i \\ a\sum t_i^2 + c\sum t_i^4 &= \sum t_i^2 y_i \\ b\sum t_i^4 + d\sum t_i^6 &= \sum t_i^3 y_i \end{aligned} \qquad (20.32)$$

Din prima şi a treia ecuaţie a sistemului (20.32) rezultă formulele de calcul simplificat ale parametrilor a şi c ai parabolei de gradul 3:

$$a = \frac{\sum t_i^4 \sum y_i - \sum t_i^2 \sum t_i^2 y_i}{n\sum t_i^4 - (\sum t_i^2)^2} \qquad (20.33)$$

$$c = \frac{n\sum t_i^2 y_i - \sum t_i^2 \sum y_i}{n\sum t_i^4 - (\sum t_i^2)^2} \qquad (20.34)$$

iar din ecuaţia a doua şi a patra rezultă:

$$b = \frac{\sum t_i^6 \sum t_i y_i - \sum t_i^4 \sum t_i^3 y_i}{\sum t_i^2 \sum t_i^6 - (\sum t_i^4)^2} \qquad (20.35)$$

$$d = \frac{\sum t_i^2 \sum t_i^3 y_i - \sum t_i^4 \sum t_i y_i}{\sum t_i^2 \sum t_i^6 - (\sum t_i^4)^2} \qquad (20.36)$$

In cazul general, tendinţa polinomială de gradul m depinde de k = m+1 parametri:

$$y_t = b_o + b_1 t + b_2 t^2 + \ldots + b_m t^m \qquad (20.37)$$

Pentru determinarea celor k parametri, considerăm k puncte de coordonate $(y_1, t_1), \ldots, (y_k, t_k)$ prin care presupunem că va trece curba y_t.

Punînd condiţia ca polinomul (20.37) să treacă prin cele k puncte, se obţine sistemul de ecuaţii:

$$y_i = b_o + b_1 t_i + \ldots + b_m t_i^m \quad (i = 1,2,\ldots,k) \qquad (20.38)$$

Transcris sub formă matricială, sistemul (20.38) devine:

$$\underbrace{\begin{Vmatrix} 1 & t_1 & \dots & t_1^m \\ 1 & t_2 & \dots & t_2^m \\ \vdots & \dots & \dots & \vdots \\ 1 & t_k & \dots & t_k^m \end{Vmatrix}}_{T} \cdot \underbrace{\begin{Vmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{Vmatrix}}_{B} = \underbrace{\begin{Vmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{Vmatrix}}_{Y} \tag{20.39}$$

ceea ce ne conduce la matricea $T_{k\times k}$, vectorul-coloană $B_{k\times 1}$ al necunoscutelor şi vectorul-coloană $Y_{k\times 1}$ al termenilor liberi.

Sintetic, relaţia (20.39) se poate scrie:

$$T_{k\times k} \cdot B_{k\times 1} = Y_{k\times 1} \tag{20.40}$$

de unde rezultă:

$$B_{k\times 1} = T^{-1}_{k\times k} \cdot Y_{k\times 1}$$

Notăm cu D determinantul matricii T:

$$D = \begin{vmatrix} 1 & t_1 & \dots & t_1^m \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 1 & t_k & \dots & t_k^m \end{vmatrix} \tag{20.41}$$

şi cu D_i $(i = 0,1,2,\dots,m)$ determinanţii obţinuţi înlocuind coloana $i+1$ din matricea T cu vectorul-coloană Y:

$$D_i = \begin{vmatrix} 1 & t_1 & \dots & t_1^{i-1} & y_1 & t_1^{i+1} & \dots & t_1^m \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ 1 & t_k & \dots & t_k^{i-1} & y_k & t_k^{i+1} & \dots & t_k^m \end{vmatrix} \tag{20.42}$$

Cu notaţiile (20.41) şi (20.42), formulele de calcul ale parametrilor b_i $(i = 0,1,\ldots,m)$ sînt, conform regulei lui Cramer:

$$b_i = \frac{D_i}{D} \qquad (i = 0,1,\ldots,m) \qquad (20.43)$$

Ecuaţiile polinomiale aproximează bine evoluţia fenomenelor în timp pe perioade determinate. Totuşi, este greu să se găsească o bază logică pentru folosirea acestor ecuaţii de trend. Acest fapt este clar ilustrat de tendinţa liniară sau de cea descrisă de un polinom de orice grad care, prin prelungire poate deveni negativă într-o direcţie sau în ambele şi va creşte sau scădea nelimitat în ambele direcţii. Este clar că cele mai multe serii dinamice reale nu pot avea valori negative şi de regulă tind spre o valoare limită superioară sau inferioară într-o direcţie sau în ambele.

Un polinom poate fi făcut să se apropie cît mai mult de evoluţia reală a unui fenomen prin creşterea gradului său. In particular, dacă polinomul are un grad egal cu numărul termenilor unei serii dinamice minus 1, atunci curba de tendinţă va trece prin toate punctele empirice, dar trendul va fi de neînţeles, deoarece numărul gradelor de libertate scade cu 1 pe măsură ce gradul polinomului creşte cu 1. Astfel, atunci cînd gradul polinomului este mai mic cu o unitate decît numărul termenilor unei serii dinamice, numărul gradelor de libertate al polinomului este egal cu zero.

Dacă se adaugă noi termeni seriilor dinamice iniţiale, operaţia de ajustare trebuie reluată de la început.

20.2.6.4. Tendinţa exponenţială

Tendinţa exponenţială este specifică fenomenelor care se dezvoltă aproximativ în progresie geometrică şi este descrisă de funcţia:

$$y = ab^t \qquad (20.44)$$

sau

$$\log y = \log a + t \log b \qquad (20.45)$$

Aplicînd metoda celor mai mici pătrate pentru estimarea parametrilor a şi b ai funcţiei (20.45), rezultă sistemul de ecua-

ţii normale:

$$n \log a + \log b \sum t_i = \sum \log y_i \qquad (20.46)$$
$$\log a \sum t_i + \log b \sum t_i^2 = \sum t_i \log y_i$$

Sistemul (20.46) se poate simplifica, dacă se introduce substituţia $\sum t_i = 0$, astfel:

$$n \log a = \sum \log y_i \qquad (20.47)$$
$$\log b \sum t_i^2 = \sum t_i \log y_i$$

Din sistemul (20.47) rezultă formulele de calcul simplificat al parametrilor ecuaţiei exponenţiale:

$$a = \text{antilog} \left(\frac{\sum \log y_i}{n} \right) \qquad (20.48)$$

$$b = \text{antilog} \left(\frac{\sum t_i \log y_i}{\sum t_i^2} \right) \qquad (20.49)$$

Pentru verificarea exactităţii calculelor de ajustare se utilzează relaţia: $\sum \log y_i = \sum \log y_{t_i}$

Considerăm, pentru exemplificarea calculului tendinţei exponenţiale, dinamica produsului social al României în perioada 1950 - 1977 (tabelul 20.8, col. 1 - 2).

Pentru a ne forma o primă imagine cu privire la forma tendinţei de evoluţie a dinamicii produsului social, se reprezintă grafic seria în fig. 20.8.

Din fig. 20.8 se observă că dinamica produsului social în perioada 1950-1977 a evoluat după modelul unei curbe exponenţiale, descrisă de ecuaţia (20.44).

Parametrii curbei (20.44) se pot determina cu ajutorul formulelor (20.48) şi (20.49). Elementele de calcul necesare aplicate acestor formule s-au determinat în coloanele 3-6 ale tabelului 20.8.

Substituind în formulele (20.48) şi (20.49) datele din tabelul 20.8, rezultă:

$$a = \text{antilog} \frac{71,7186}{28} = \text{antilog } 2,5614 = 364,237$$

şi

$$b = \text{antilog} \frac{139,4458}{7308} = \text{antilog } 0,0191 = 1,049$$

Tabelul 20.8. Dinamica produsului social al României în perioada 1950-1977

Anii	y_i (%)	t_i	t_i^2	log y_i	t_i log y_i	log y_{t_i}	y_{t_i}
1	2	3	4	5	6	7	8
1950	100	-27	729	2,0000	-54,000	2,0462	111,22
1951	125	-25	625	2,0969	-52,4227	2,0843	121,44
1952	137	-23	529	2,1367	-49,1446	2,1225	132,59
1953	157	-21	441	2,1959	-46,1139	2,1607	144,77
1954	161	-19	361	2,2068	-41,9297	2,1988	158,06
1955	186	-17	289	2,2695	-38,5817	2,2370	172,58
1956	187	-15	225	2,2718	-34,0776	2,2752	188,43
1957	207	-13	169	2,3160	-30,1076	2,3133	205,74
1958	215	-11	121	2,3324	-25,6568	2,3515	224,64
1959	237	-9	81	2,3747	-21,3727	2,3896	245,27
1960	263	-7	49	2,4200	-16,9397	2,4278	267,80
1961	293	-5	25	2,4669	-12,3343	2,4660	292,40
1962	311	-3	9	2,4928	-7,4783	2,5041	319,25
1963	338	-1	1	2,5289	-2,5289	2,5423	348,57
1964	377	1	1	2,5763	2,5763	2,5805	380,59
1965	414	3	9	2,6170	7,8510	2,6186	415,55
1966	457	5	25	2,6599	13,2996	2,6568	453,72
1967	501	7	49	2,6998	18,8989	2,6949	495,39
1968	540	9	81	2,7324	24,5915	2,7331	540,89
1969	582	11	121	2,7649	30,4142	2,7713	590,57
1970	629	13	169	2,7987	36,3825	2,8094	644,81
1971	701	15	225	2,8457	42,6858	2,8476	704,04
1972	771	17	289	2,8871	49,0799	2,8858	768,70
1973	852	19	361	2,9304	55,6783	2,9239	839,31
1974	947	21	441	2,9763	62,5033	2,9621	916,40
1975	1000	23	529	3,0000	69,0000	3,0002	1000,56
1976	1100	25	625	3,0414	76,0348	3,0384	1092,46
1977	1200	27	729	3,0792	83,1379	3,0766	1192,81
Total	-	0	7308	71,7186	139,4458	71,7186	-

Deci tendința de evoluție a dinamicii produsului social în perioada 1950-1977 este descrisă de modelul:

$$y_{t_i} = 364,237 \cdot 1,049^{t_i}$$

sau, în formă liniarizată:

$$\log y_{t_i} = 2,5614 + 0,0191\ t_i$$

Tabelul 20.8. Dinamica produsului social al României în perioada 1950-1977

Anii	y_i (%)	t_i	t_i^2	log y_i	t_i log y_i	log y_{t_i}	y_{t_i}
1	2	3	4	5	6	7	8
1950	100	-27	729	2,0000	-54,000	2,0462	111,22
1951	125	-25	625	2,0969	-52,4227	2,0843	121,44
1952	137	-23	529	2,1367	-49,1446	2,1225	132,59
1953	157	-21	441	2,1959	-46,1139	2,1607	144,77
1954	161	-19	361	2,2068	-41,9297	2,1988	158,06
1955	186	-17	289	2,2695	-38,5817	2,2370	172,58
1956	187	-15	225	2,2718	-34,0776	2,2752	188,43
1957	207	-13	169	2,3160	-30,1076	2,3133	205,74
1958	215	-11	121	2,3324	-25,6568	2,3515	224,64
1959	237	-9	81	2,3747	-21,3727	2,3896	245,27
1960	263	-7	49	2,4200	-16,9397	2,4278	267,80
1961	293	-5	25	2,4669	-12,3343	2,4660	292,40
1962	311	-3	9	2,4928	-7,4783	2,5041	319,25
1963	338	-1	1	2,5289	-2,5289	2,5423	348,57
1964	377	1	1	2,5763	2,5763	2,5805	380,59
1965	414	3	9	2,6170	7,8510	2,6186	415,55
1966	457	5	25	2,6599	13,2996	2,6568	453,72
1967	501	7	49	2,6998	18,8989	2,6949	495,39
1968	540	9	81	2,7324	24,5915	2,7331	540,89
1969	582	11	121	2,7649	30,4142	2,7713	590,57
1970	629	13	169	2,7987	36,3825	2,8094	644,81
1971	701	15	225	2,8457	42,6858	2,8476	704,04
1972	771	17	289	2,8871	49,0799	2,8858	768,70
1973	852	19	361	2,9304	55,6783	2,9239	839,31
1974	947	21	441	2,9763	62,5033	2,9621	916,40
1975	1000	23	529	3,0000	69,0000	3,0002	1000,56
1976	1100	25	625	3,0414	76,0348	3,0384	1092,46
1977	1200	27	729	3,0792	83,1379	3,0766	1192,81
Total	-	0	7308	71,7186	139,4458	71,7186	-

Deci tendința de evoluție a dinamicii produsului social în perioada 1950-1977 este descrisă de modelul:

$$y_{t_i} = 364,237 \cdot 1,049^{t_i}$$

sau, în formă liniarizată:

$$\log y_{t_i} = 2,5614 + 0,0191\ t_i$$

Substituind succesiv valorile lui t_i din col. 3 a tabelului 20.8 în aceste modele, s-au obţinut valorile teoretice ale dinamicii produsului social în coloanele 7 şi 8 ale aceluiaşi tabel. Din tabel se observă că:

$$\sum \log y_i = \sum \log y_{t_i} = 71,7186$$

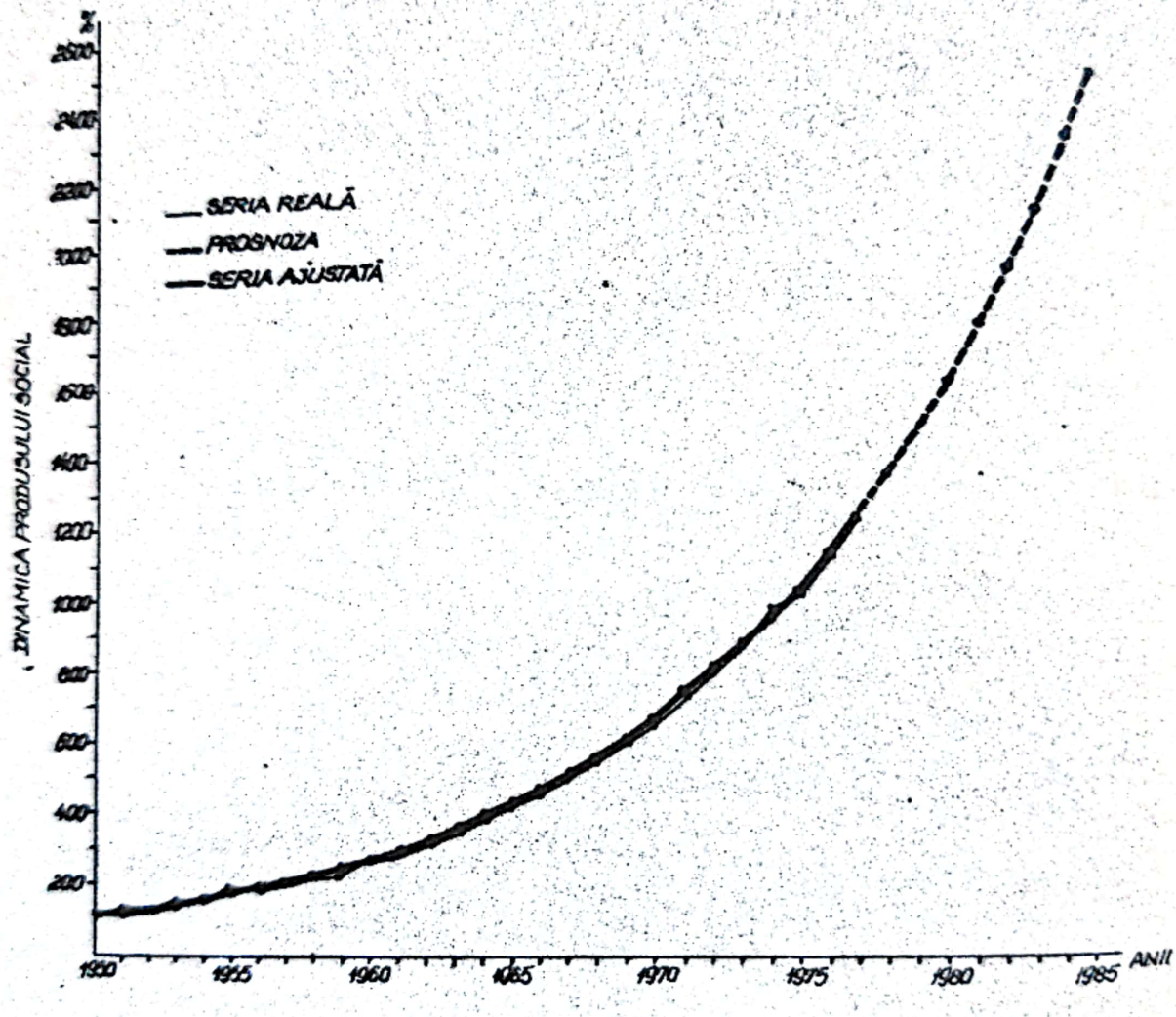


Fig. 20.8. DINAMICA PRODUSULUI SOCIAL AL ROMÂNIEI ÎN PERIOADA 1950-1977 ŞI PROGNOZA PÎNĂ ÎN ANUL 1985

Atît din tabelul 20.8, cît şi din fig. 20.8 se observă că funcţia exponenţială descrie în mod corespunzător evoluţia dinamicii produsului social din România. Datorită acestui fapt, este logic ca prognoza dinamicii produsului social pentru perioada imediat următoare să se facă pe baza modelului exponenţial ob-

ţinut pentru perioada 1950 - 1977. Introducînd în model valorile variabilei t_i aferente perioadei 1978-1985 rezultă:

Anii	Dinamica produsului social (%)
1978	1302,35
1979	1421,97
1980	1552,57
1981	1695,17
1982	1850,87
1983	2020,86
1984	2206,48
1985	2409,14

Prognoza dezvoltării fenomenelor sociale economice după modelul curbei exponenţiale trebuie să se facă cu multă prudenţă şi pentru orizonturi nu prea îndepărtate de timp. Chiar dacă curba exponenţială este potrivită pentru ajustarea unei serii dinamice, extrapolarea ei pentru orizonturi mari de timp duce la rezultate neverosimile.

20.2.6.5. Ajustarea după o curbă exponenţială modificată

Tendinţele de dezvoltare ale fenomenelor economice, care nu prezintă ritmuri de variaţie cu bază mobilă apropiate ca mărime se pot determina cu ajutorul curbei exponenţiale modificate. Modificarea constă în adunarea algebrică a unei constante k termenilor iniţiali ai seriei dinamice, obţinîndu-se în acest fel o nouă serie dinamică, apropiată mai mult de modelul unei progresii geometrice:

$$y = k + a.b^t \qquad (20.50)$$

sau

$$y - k = a.b^t \qquad (20.51)$$

care se mai poate scrie:

$$\frac{y-k}{a} = e^{t \ln b} \qquad (20.52)$$

Dacă se aplică metoda celor mai mici pătrate pentru determinarea parametrilor curbei exponenţiale modificate, se obţine următorul sistem de ecuaţii normale:

$$\begin{aligned} nk + a\sum b^{t_i} &= \sum y_i \\ k\sum b^{t_i} + a\sum b^{2t_i} &= \sum b^{t_i} y_i \\ k\sum t_i b^{t_i-1} + a\sum t_i b^{2t_i-1} &= \sum t_i y_i b^{t_i-1} \end{aligned} \qquad (20.53)$$

Sistemul obţinut este complicat şi greu de rezolvat, de aceea se recurge în practică la alte metode de estimare a parametrilor unei curbe exponenţiale modificate. Cea mai des utilizată este metoda punctelor selecţionate şi o variantă a acesteia - metoda mediilor selecţionate. Prezentăm, în continuare, metoda punctelor selecţionate.

Logaritmînd ambii membri al ecuaţiei (20.52), rezultă:

$$\ln \frac{y - k}{a} = t \ln b \qquad (20.54)$$

Pentru determinarea constantei k considerăm trei puncte echidistante, simetrice faţă de origine $(-t_1, y_1)$, $(0, y_2)$, (t_1, y_3) şi, punînd condiţia ca ecuaţia (20.54) să treacă prin aceste puncte, se obţine următorul sistem de ecuaţii:

$$\begin{aligned} \ln \frac{y_1 - k}{a} &= -t_i \ln b \\ \ln \frac{y_2 - k}{a} &= 0 \\ \ln \frac{y_3 - k}{a} &= t_1 \ln b \end{aligned} \qquad (20.55)$$

Din ecuaţia a doua a sistemului (20.54) rezultă:

$$\frac{y_2 - k}{a} = 1 \quad \text{sau} \quad y_2 - k = a \qquad (20.56)$$

Din prima şi a treia ecuaţie a sistemului (20.55), prin însumare, obţinem:

$$\ln \frac{y_1 - k}{a} + \ln \frac{y_3 - k}{a} = 0, \text{ adică } \frac{y_1 - k}{a} \cdot \frac{y_3 - k}{a} = 1 \qquad (20.57)$$

Efectuînd calculele în (20.57) se obţine:

$$(y_1-k).(y_3-k) = a^2 \quad \text{sau}$$

$$y_1y_3-ky_1-ky_3+k^2 = a^2 \qquad (20.58)$$

Ţinînd seama de relaţia (20.56), ecuaţia (20.58) devine:

$$y_1y_3-ky_1-ky_3+k^2 = y_2^2-2ky_2+k^2 \quad \text{sau}$$

$$y_1y_3-y_2^2 = ky_1+ky_3-2ky_2 \quad \text{de unde}$$

$$k = \frac{y_1y_3 - y_2^2}{y_1+y_3-2y_2} \qquad (20.59)$$

In formula (20.59) valorile y_1, y_2 şi y_3 sînt fie termeni ai seriei, fie medii formate pentru fiecare treime a ei.

Ajustarea unei serii după modelul curbei exponenţiale modificate presupune parcurgerea următoarelor etape de calcul:

1) Se împarte seria iniţială în trei părţi egale şi apoi se determină media fiecărei grupe de termeni, obţinîndu-se y_1, y_2 şi y_3. Dacă numărul termenilor seriei dinamice nu este divizibil cu 3, se pot considera intervale care se suprapun unul peste altul, cu condiţia ca distanţele dintre y_1, y_2 şi y_3 să fie egale între ele.

2) Se determină valoarea constantei k pe baza formulei (20.59).

3) Se determină termenii seriei empirice modificate prin adunarea algebrică a valorii k la termenii reali ai seriei date.

4) Se ajustează seria modificată după o funcţie exponenţială de forma $y' = y - k = a.b^t$.

5) Se recalculează valorile tendinţei prin scăderea algebrică a constantei k.

Pentru exemplificare să considerăm exemplul anterior (tabelul 20.8) referitor la dinamica produsului social şi să ajustăm această serie dinamică după modelul unei curbe exponenţiale modificate. Deoarece numărul termenilor seriei nu este divizibil cu 3, renunţăm la primul termen. Cele trei puncte selecţionate vor avea coordonatele:

t_i	y_i
1951	125
1964	377
1977	1200

Substituind aceste date în formula (20.59) se obţine:

$$k = \frac{125 \cdot 1200 - 377^2}{125 + 1200 - 2 \cdot 377} = 13{,}785$$

Aceasta înseamnă că dinamica reală a produsului social se abate de la modelul unei progresii geometrice în plus cu k = 13,785%. Scăzînd constanta k din termenii reali ai seriei se obţine seria modificată (col. 3 a tabelului 20.9).

Tabelul 20.9. Ajustarea seriei dinamice după modelul curbei exponenţiale modificate

Anii	y_i (q)	$y'_i = y_i - k$	$\log y'_i$	t_i	t_i^2	$t_i \log y'_i$	$\log y'_{t_i}$	y'_{t_i}	y_{t_i}
1	2	3	4	5	6	7	8	9	10
1951	125	111	2,0462	-13	169	-26,6001	2,0500	112	126
1952	137	123	2,0907	-12	144	-25,0880	2,0894	123	137
1953	157	143	2,1560	-11	121	-23,7159	2,1288	135	148
1954	161	147	2,1680	-10	100	-21,6795	2,1682	147	161
1955	186	172	2,2361	-9	81	-20,1246	2,2076	161	175
1956	187	173	2,2386	-8	64	-17,9087	2,2470	177	190
1957	207	193	2,2860	-7	49	-16,0023	2,2863	193	207
1958	215	201	2,3037	-6	36	-13,8220	2,3257	212	225
1959	237	223	2,3487	-5	25	-11,7436	2,3651	232	246
1960	263	249	2,3966	-4	16	-9,5863	2,4045	254	268
1961	293	279	2,4459	-3	9	-7,3378	2,4439	278	292
1962	311	297	2,4731	-2	4	-4,9461	2,4833	304	318
1963	338	324	2,5108	-1	1	-2,5108	2,5227	333	347
1964	377	363	2,5602	0	0	0,0000	2,5621	365	379
1965	414	400	2,6023	1	1	2,6023	2,6014	399	413
1966	457	443	2,6466	2	4	5,2932	2,6408	437	451
1967	501	487	2,6877	3	9	8,0632	2,6802	479	493
1968	540	526	2,7212	4	16	10,8847	2,7196	524	538
1969	582	568	2,7545	5	25	13,7726	2,7590	574	588
1970	629	615	2,7890	6	36	16,7342	2,7984	629	642
1971	701	687	2,8371	7	49	19,8596	2,8378	688	702
1972	771	757	2,8792	8	64	23,0338	2,8772	754	767
1973	852	838	2,9234	9	81	26,3102	2,9165	825	839
1974	947	933	2,9700	10	100	29,6998	2,9559	904	917
1975	1000	986	2,9940	11	121	32,9337	2,9953	989	1003
1976	1100	1086	3,0359	12	144	36,4310	3,0347	1083	1097
1977	1200	1186	3,0742	13	169	39,9641	3,0741	1186	1200
Total	-	-	69,1755	0	1638	64,5164	69,1755	-	-

După ce s-au determinat valorile transformate y'_i se ajustează seria astfel modificată după modelul curbei exponențiale (20.51). Calculele intermediare se prezintă în tabelul 20.9.

Valorile parametrilor curbei exponențiale modificate se determină tot pe baza formulelor (20.48) și (20.49), astfel:

$$a = \text{antilog } \frac{69,1755}{27} = \text{antilog } 2,56206 = 364,800$$

și

$$b = \text{antilog } \frac{64,5164}{1638} = \text{antilog } 0,03939 = 1,095$$

Deci, curba exponențială modificată va fi:

$$y'_{t_i} = 364,8 \cdot 1,095^{t_i}$$

sau

$$\log y'_{t_i} = 2,56206 + 0,03939\ t_i$$

Substituind succesiv în aceste formule valorile lui t_i, se vor obține valorile seriei modificate ajustate din coloanele 8 și 9 ale tabelului 20.9.

Considerînd că și în următorii ani dinamica produsului social va evolua după același model ca și în perioada 1951-1977, s-a calculat prognoza pînă în anul 1985 (tabelul 20.10).

Tabelul 20.10. Prognoza dinamicii produsului social în perioada 1978 - 1985

Anii	Dinamica produsului social (%)
1978	1298
1979	1421
1980	1556
1981	1704
1982	1866
1983	2043
1984	2237
1985	2450

Parametrii unei curbe exponențiale modificate se pot determina și cu ajutorul metodei mediilor selecționate, mai simplu de calculat și cu posibilități de generalizare și pentru curba lui Gompertz și curba logistică.

Pentru aceasta se împarte seria în trei părţi egale, iar dacă numărul seriei nu este divizibil cu 3, atunci se neglijează unul sau doi termeni de la începutul seriei.

Considerăm o serie formată din 15 termeni şi împărţită în trei grupe de câte 5 termeni (tabelul 20.11).

Tabelul 20.11. Metoda mediilor selecţionate

Anii (t_i)	y_i	$k + ab^t$	Mediile perioadelor
0	y_0	$k + ab^0$	
1	y_1	$k + ab^1$	
2	y_2	$k + ab^2$	$\bar{y}_1$
3	y_3	$k + ab^3$	
4	y_4	$k + ab^4$	
Total	$\sum_0^4 y_i$	$5k + a\sum_0^4 b^{t_i}$	
5	y_5	$k + ab^5$	
6	y_6	$k + ab^6$	
7	y_7	$k + ab^7$	$\bar{y}_2$
8	y_8	$k + ab^8$	
9	y_9	$k + ab^9$	
Total	$\sum_5^9 y_i$	$5k + a\sum_5^9 b^{t_i}$	
10	y_{10}	$k + ab^{10}$	
11	y_{11}	$k + ab^{11}$	
12	y_{12}	$k + ab^{12}$	$\bar{y}_3$
13	y_{13}	$k + ab^{13}$	
14	y_{14}	$k + ab^{14}$	
Total	$\sum_{10}^{14} y_i$	$5k + a\sum_{10}^{14} b^{t_i}$	

Scanned with OKEN Scanner

Mediile celor trei perioade se determină astfel:

$$\bar{y}_1 = \frac{\sum_{0}^{4}(k + ab^{t_i})}{5} = k + \frac{a}{5} \cdot \sum_{0}^{4} b^{t_i}$$

$$\bar{y}_2 = k + \frac{a}{5} \cdot \sum_{5}^{9} b^{t_i}$$

$$\bar{y}_3 = k + \frac{a}{5} \cdot \sum_{10}^{14} b^{t_i}$$

Sau, generalizînd pentru orice număr "n" de termeni cuprinşi într-o grupă:

$$\bar{y}_1 = k + a \cdot \frac{\sum_{0}^{n-1} b^{t_i}}{n} \qquad (20.60)$$

$$\bar{y}_2 = k + a \cdot \frac{\sum_{n}^{2n-1} b^{t_i}}{n} \qquad (20.61)$$

$$\bar{y}_3 = k + a \cdot \frac{\sum_{2n}^{3n-1} b^{t_i}}{n} \qquad (20.62)$$

Scăzînd pe (20.60) din (20.61) rezultă:

$$\bar{y}_2 - \bar{y}_1 = a \cdot \frac{\sum_{n}^{2n-1} b^{t_i} - \sum_{0}^{n-1} b^{t_i}}{n} \qquad (20.63)$$

şi pe (20.61) din (20.62) deducem:

$$\bar{y}_3 - \bar{y}_2 = a \cdot \frac{\sum_{2n}^{3n-1} b^{t_i} - \sum_{n}^{2n-1} b^{t_i}}{n} \qquad (20.64)$$

Raportînd pe (20.64) la (20.63), obţinem:

$$\frac{\bar{y}_3 - \bar{y}_2}{\bar{y}_2 - \bar{y}_1} = \frac{\sum_{2n}^{3n-1} b^{t_i} - \sum_{n}^{2n-1} b^{t_i}}{\sum_{n}^{2n-1} b^{t_i} - \sum_{o}^{n-1} b^{t_i}} = b^n \qquad (20.65)$$

de unde:

$$b = \sqrt[n]{\frac{\bar{y}_3 - \bar{y}_2}{\bar{y}_2 - \bar{y}_1}} \qquad (20.66)$$

Valoarea parametrului a rezultă din relaţia (20.63) sau din (20.64):

$$a = \frac{n.(\bar{y}_2 - \bar{y}_1}{\sum_{n}^{2n-1} b^{t_i} - \sum_{o}^{n-1} b^{t_i}} \qquad (20.67)$$

Substituind acum pe a şi b în una dintre relaţiile (20.60), (20.61) sau (20.62), rezultă valoarea constantei k :

$$k = \bar{y}_1 - a \cdot \frac{\sum_{o}^{n-1} b^{t_i}}{n} \qquad (20.68)$$

Aplicarea practică a metodei mediilor selecţionate o vom efectua pe baza seriei din tabelul 20.8. Rezultatele intermediare şi cele finale sînt prezentate în tabelul 20.12.

$$b = \sqrt[9]{\frac{864,666 - 388,222}{388,222 - 179,111}} = \sqrt[9]{2,278}$$

$$\log b = \frac{\log 2,278}{9} = \frac{0,35755}{9} = 0,03973$$

de unde rezultă:

$$b = 1,096$$

$$a = \frac{9.(388,222 - 179,111)}{30,464 - 13,350} = \frac{1881,999}{17,111} = 109,987$$

şi

$$k = 179,111 - 109,987 \cdot \frac{13,350}{9} = 15,964$$

Tabelul 20.12. Ajustarea seriei dinamice prin metoda mediilor selectionate

Anii	Dinamica produsului social (%)	t_i	Mediile selectionate	b^{t_i}	ab^{t_i}	$y_{t_i}=k+ab^{t_i}$
1	2	3	4	5	6	7
1951	125	0		1,000	109,987	125,951
1952	137	1		1,096	120,545	136,510
1953	157	2		1,201	132,094	148,058
1954	161	3		1,316	144,743	160,707
1955	186	4	$\bar{y}_1 = 179,111$	1,442	158,601	174,565
1956	187	5		1,581	173,889	189,853
1957	207	6		1,733	190,607	206,571
1958	215	7		1,899	208,865	224,829
1959	237	8		2,082	228,993	244,957
Total 1	1612	-		13,350		
1960	263	9		2,281	250,880	266,844
1961	293	10		2,500	274,968	290,932
1962	311	11		2,741	301,474	317,438
1963	338	12		3,004	330,401	346,365
1964	377	13	$\bar{y}_2 = 388,222$	3,292	362,077	378,041
1965	414	14		3,608	396,833	412,797
1966	457	15		3,954	434,888	450,853
1967	501	16		4,334	476,684	492,647
1968	540	17		4,750	522,438	538,402
Total 2	3494	-		30,464		
1969	582	18		5,206	572,592	588,556
1970	629	19		5,706	627,586	643,550
1971	701	20		6,254	687,859	703,823
1972	771	21		6,854	753,851	769,815
1973	852	22	$\bar{y}_3 = 864,666$	7,512	826,222	842,186
1974	947	23		8,233	905,523	921,487
1975	1000	24		9,023	992,413	1008,377
1976	1100	25		9,890	1087,771	1103,735
1977	1200	26		10,839	1192,149	1208,113
Total 3	7782	-		69,517		

Inlocuind cu valorile lor pe a, b si k in relatia (20.50) rezultă:

$$y_{t_i} = 15,964 + 109,987 \cdot 1,096^{t_i}$$

Substituind în această ecuaţie valorile lui t_i din col.3 a tabelului 20.12, rezultă valorile seriei ajustate din col. 7.

20.2.6. Curba lui Gompertz

Această curbă a fost introdusă în anul 1825 de statisticianul englez Benjamin Gompertz[1] (1779-1865) şi aplicată pe scară largă mai ales în demografie pentru extrapolarea probabilităţilor de deces la vîrstele înaintate; mai recent, ea se aplică în modelarea creşterii economice şi marketing.

Curba lui Gompertz are forma unui S alungit şi descrie evoluţia unui fenomen ale căror sporuri logaritmice de creştere se micşorează continuu cu o raţie constantă. Formula obişnuită de calcul este descrisă de funcţia:

$$y = a \cdot b^{c^t} \qquad (20.69)$$

Curba lui Gompertz poate fi scrisă într-o formă similară curbei exponenţiale modificate, prin logaritmarea expresiei (20.69), astfel:

$$\log y = \log a + c^t \cdot \log b \qquad (20.70)$$

Primul termen din membrul a doilea al relaţiei (20.70), adică $\log a$, reprezintă valoarea maximă sau "plafonul" către care tinde curba lui Gompertz, iar termenul al doilea reprezintă distanţa care separă, la un moment dat, nivelul atins de fenomenul luat calcul de nivelul maxim. Desigur, această distanţă se micşorează din ce în ce mai mult pe măsură ce se înaintează în timp. Parametrul c are valorile cuprinse între zero şi unu:

$$0 \leq c \leq 1$$

Din relaţia (20.70) se observă că dacă se face substituţia $\log y = y'$, $\log a = a'$ şi $\log b = b'$, se obţine ecuaţia unei curbe exponenţiale modificate:

---

1) Gompertz B., "On the nature of the function expressive of the law of human mortality", în: "Philosophical Transactions of the Royal Society of London", 1825.

$$y' = a' + b'c^t \qquad (20.71)$$

Curba lui Gompertz are proprietatea că primele diferenţe ale valorilor $\log y_i$ au o raţie constantă de creştere şi egală cu b. In cazul general, valoarea lui a este negativă, iar valoarea parametrului b este cuprinsă între zero şi unu.

Dacă se reprezintă pe scară semilogaritmică ($\log y_i$ pe ordonată şi t_i pe abscisă), atunci funcţia (20.71) ia o formă de curbă concavă, ceea ce înseamnă că rata de creştere devine din ce în ce mai mică pe măsură ce t creşte.

Reprezentată într-un sistem de axe rectangulare pe scară aritmetică, curba lui Gompertz ia forma unui S asimetric şi are două asimptote: cea inferioară egală cu zero, iar cea superioară egală cu a'. Asimetria acestei curbe se explică prin comportamentul diferit al ei, de o parte şi de alta a punctului de inflexiune.

Curba lui Gompertz poate fi scrisă şi sub formă liniară:

$$\log \log \left(\frac{y}{a}\right) = A + Bx \qquad (20.72)$$

unde $A = \log b'$ şi $B = c$.

Dacă în locul logaritmilor zecimali se vor folosi logaritmi naturali în ecuaţia (20.72), atunci nu se schimbă valorile lui y şi ale lui a, ci doar cele ale lui b şi c.

Ca şi în cazul curbei exponenţiale modificate, pentru determinarea parametrilor curbei lui Gompertz nu se poate utiliza metoda celor mai mici pătrate, care ar duce la un sistem de ecuaţii complicat şi deci greu de rezolvat. De aceea, se recomandă utilizarea în practică a metodei mediilor selecţionate.

In cazul curbei lui Gompertz aplicarea metodei mediilor selecţionate pentru determinarea parametrilor se face în mod asemănător cu cazul curbei exponenţiale modificate, dar numai după liniarizarea pe baza formulei (20.70).

Parametrii curbei Gompertz vor avea următoarele expresii:

$$c = \sqrt[n]{\frac{\sum_{2n}^{3n-1} \log y_i - \sum_{n}^{2n-1} \log y_i}{\sum_{n}^{2n-1} \log y_i - \sum_{0}^{n-1} \log y_i}} \qquad (20.73)$$

$$\log b = \frac{\sum_{n}^{2n-1} \log y_i - \sum_{0}^{n-1} \log y_i}{\sum_{n}^{2n-1} c^{t_i} - \sum_{0}^{n-1} c^{t_i}} \qquad (20.74)$$

$$\log a = \frac{\sum_{0}^{n-1} \log y_i - \log b \sum_{0}^{n-1} c^{t_i}}{n} \qquad (20.75)$$

Elementele de calcul necesare aplicării acestor formule se prezintă sistematizat în tabelul 20.13.

Tabelul 20.13. Aplicarea metodei mediilor selecţionate în cazul curbei lui Gompertz

Anii (t_i)	y_i	$\log y_i$	Mediile selecţionate
0 1 ⋮ n-1	y_0 y_1 y_{n-1}	$\log y_0$ $\log y_1$ $\log y_{n-1}$	$\frac{\sum_{0}^{n-1} \log y_i}{n}$
Total 1		$\sum_{0}^{n-1} \log y_i$	
n n+1 ⋮ 2n-1	y_n y_{n+1} y_{2n-1}	$\log y_n$ $\log y_{n+1}$ $\log y_{2n-1}$	$\frac{\sum_{n}^{2n-1} \log y_i}{n}$
Total 2		$\sum_{n}^{2n-1} \log y_i$	

Tabelul 20.13 (continuare)

2n 2n+1 ⋮ 3n-1	y_{2n} y_{2n+1} y_{3n-1}	$\log y_{2n}$ $\log y_{2n+1}$ $\log y_{3n-1}$	$\sum_{2n}^{3n-1} \log y_i$
Total 3		$\sum_{2n}^{3n-1} \log y_i$	

Să considerăm, pentru exemplificarea calculului parametrilor curbei lui Gompertz, seria dinamică din tabelul 20.14, cu privire la evoluţia producţiei de oţel a României în perioada 1950-1976. Reprezentînd grafic seria (fig. 20.9) se observă că ea se apropie de forma unei curbe Gomperts, depăşind puţin punctul ei de inflexiune. Calculele preliminare necesare pentru determinarea parametrilor a, b şi c se prezintă în tabelul 20.14.

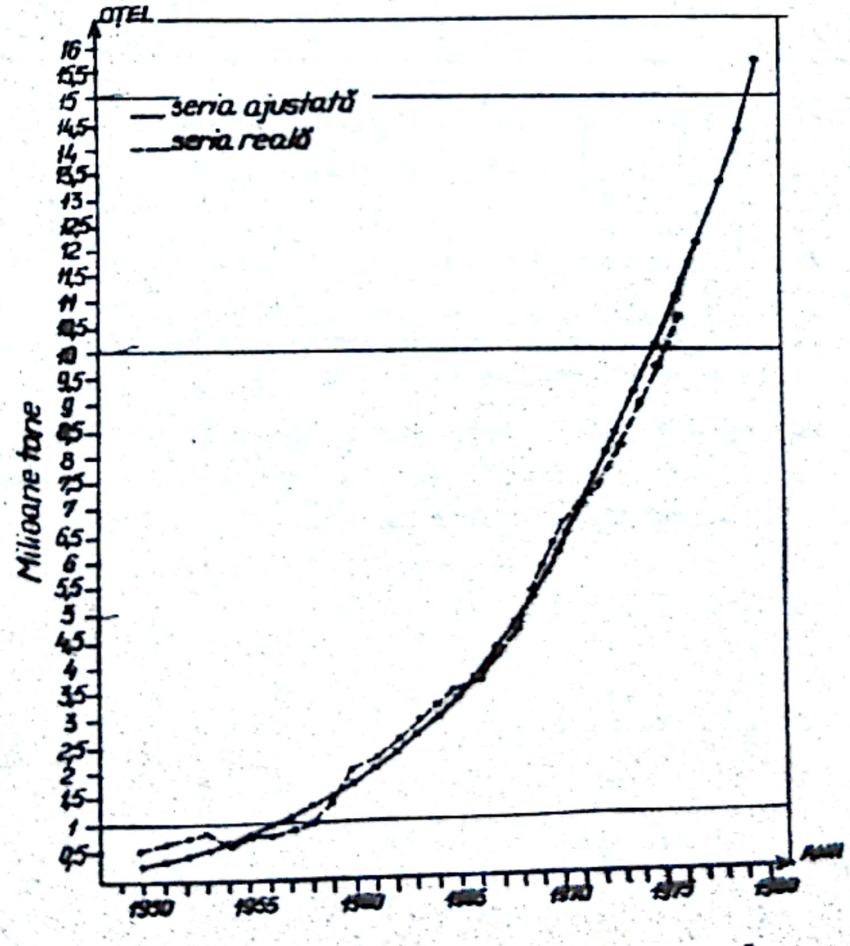


Fig.20.9. EVOLUŢIA PRODUCŢIEI DE OŢEL A ROMÂNIEI ÎN PERIOADA 1950-1976 ŞI PROGNOZA PÎNĂ ÎN ANUL 1980

Tabelul 20.14. Determinarea curbei Gompertz de evoluţie a producţiei de oţel a României în perioada 1950 - 1976

Anii	Oţel (mii tone) y_i	$\log y_i$	t_i	c^{t_i}	$c^{t_i}.\log b$	$\log y_{t_i}$	y_{t_i}
1	2	3	4	5	6	7	8
1950	555	2,74429	0	1,00000	-3,18406	2,57654	377
1951	644	2,80899	1	0,97657	-3,10945	2,65116	448
1952	698	2,84385	2	0,95368	-3,03658	3,72402	530
1953	717	2,85552	3	0,93133	-2,96542	2,79519	624
1954	628	2,79796	4	0,90950	-2,89592	2,86468	732
1955	766	2,88423	5	0,88819	-2,82806	2,93255	856
1956	779	2,89154	6	0,86738	-2,76178	2,99882	997
1957	864	2,93651	7	0,84705	-2,69705	3,06355	1158
1958	934	2,97035	8	0,82720	-2,63385	3,12675	1339
1959	1420	3,15229	9	0,80781	-2,57213	3,18847	1543
1960	1806	3,25672	10	0,78888	-2,51185	3,24875	1773
1961	2126	3,32756	11	0,77039	-2,45298	3,30762	2031
1962	2451	3,38934	12	0,75234	-2,39550	3,36510	2318
1963	2704	3,43201	13	0,73471	-2,33936	3,42124	2638
1964	3039	3,48273	14	0,71749	-2,28454	3,47606	2993
1965	3426	3,53479	15	0,70068	-2,23100	3,52960	3385
1966	3670	3,56467	16	0,68426	-2,17872	3,58189	3818
1967	4088	3,61151	17	0,66822	-2,12766	3,63294	4295
1968	4751	3,67678	18	0,65256	-2,07780	3,68281	4817
1969	5540	3,74351	19	0,63727	-2,02910	3,73150	5389
1970	6517	3,81405	20	0,62233	-1,98155	3,77905	6012
1971	6803	3,83270	21	0,60775	-1,93511	3,82549	6691
1972	7401	3,86929	22	0,59351	-1,88977	3,87084	7427
1973	8161	3,91174	23	0,57960	-1,84548	3,91512	8225
1974	8848	3,94685	24	0,56602	-1,80223	3,95837	9086
1975	9549	3,97996	25	0,55275	-1,76000	4,00061	10014
1976	10733	4,03072	26	0,53980	-1,71875	4,04185	11012

Sursa: Anuarele statistice ale R.S. România

S-a împărţit seria în trei părţi egale formate din câte 9 termeni şi apoi s-au determinat sumele logaritmilor valorilor lui y pentru fiecare parte, obţinîndu-se:

$$\sum_{0}^{n-1} \log y_i = 25,73314$$

$$\sum_{n}^{2n-1} \log y_i = 30,75162$$

$$\sum_{2n}^{3n-1} \log y_i = 34,80560$$

Substituind aceste valori în formulele (20.72), (20.73) şi (20.74) rezultă:

$c = 0,97657$

$\log b = -3,18406$

$\log a = 5,76060$

Aceasta înseamnă că evoluţia producţiei de oţel a României în perioada 1950 - 1976 poate fi descrisă de următoarea ecuaţie Gompertz:

$$\log y_t = 5,76060 + 0,97657^{t_i} \cdot (-3,18406)$$

Substituind succesiv valorile lui t_i în ecuaţia de mai sus, rezultă valorile ajustate ale producţiei de oţel, astfel:

y_{1950} = antilog (5,76060-1.3,18406) = 377 mii tone

y_{1951} = antilog (5,76060-0,97657.3,18406) = 448 mii tone

. . . . . .

y_{1976} = antilog (5,76060-0,53980.3,18406) =11012 mii tone

Valorile seriei ajustate, astfel calculate, s-au trecut în col. 8 a tabelului 20.14.

Nivelurile viitoare probabile se pot determina pe baza ecuaţiei trendului aferentă perioadei 1950 - 1976.

In anul 1977 avem $t_i = 27$ iar $c^{27} = c^{26}.c = 0,53980. 0,97657 = 0,52715$, deci:

y_{1977} = antilog(5,76060-0,52715.3,18406)=12082 mii tone

şi, procedînd similar:

y_{1978} = 13227 mii tone

y_{1979} = 14450 mii tone

y_{1980} = 15754 mii tone etc.

20.2.6.7. Curba logistică

Curba logistică a fost introdusă în anul 1845 de către matematicianul belgian P.F. Verhulst [1] şi a fost utilizată la început în domeniul biologiei şi demografiei şi mai tîrziu în econometrie. Se pare că prima încercare de aplicare în economie ar fi fost efectuată de S. Kuznets în anul 1930 în lucrarea "Secular Movement in Production and Prices".

In concepţia autorului ei, curba logistică se poate utiliza pentru a exprima caracterul legic al evoluţiei neuniforme a fenomenelor care trec întotdeauna prin două faze succesive diferite: prima fază se caracterizează printr-un ritm accelerat, iar a doua printr-un ritm descrescător, care în final tinde spre zero. Dezvoltarea unui fenomen după o curbă logistică are loc între două limite, una inferioară, care se confundă cu axa absciselor ("minimul absolut" după Verhulst) şi una superioară ("maximul absolut") care se stabileşte prin ajustare.

Dacă notăm ritmul de creştere pe unitate de timp $\frac{\Delta y}{\Delta t}$, acesta este la început crescător pînă atinge un maximum, apoi în partea a doua a intervalului de timp, începe să descrească pînă cînd devine egal cu zero (fig. 20.10).

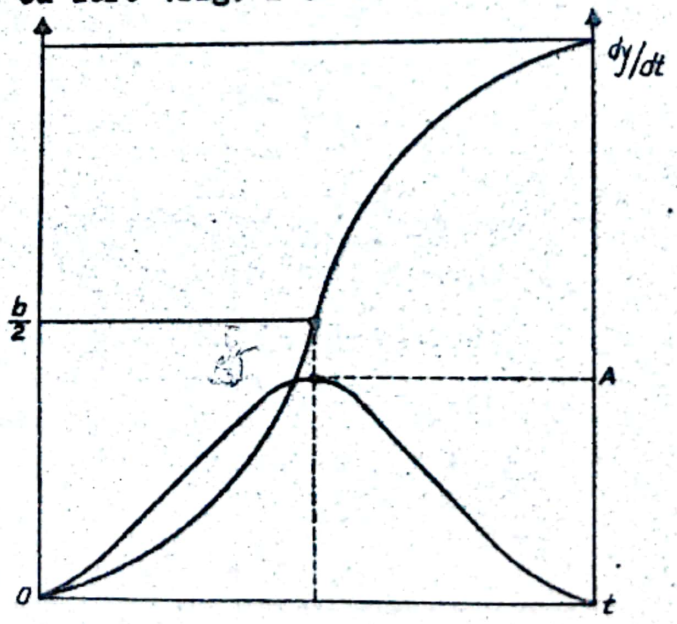


Fig. 20.10. CURBA LOGISTICĂ ŞI RITMUL DE CREŞTERE A UNEI VARIABILE LOGISTICE

1) Verhulst P.F. "Recherche sur la loi de l'accroissement de la population", în "Nouveaux mémoires de l'académie Royale des Sciences et Belles Lettres de Bruxelles", XVIII, 1845, pp 1-38.

Creşterea instantanee a ritmului de creştere se poate exprima printr-o parabolă ce trece prin origine:

$$f'(y) = \frac{dy}{dt} = - Ay^2 + B_y \qquad (20.76)$$

unde parametriii A şi B sînt pozitivi, iar curba atinge un punct de maxim cînd $f''(y)$ este nulă:

$$f''(y) = -2Ay + B = 0 \quad , \text{ de unde} \qquad (20.77)$$

$$y = \frac{B}{2A} \quad , \text{ sau notînd } \frac{B}{A} = b \text{ , rezultă } y = \frac{b}{2}$$

Ecuaţia (20.77) reprezintă o dreaptă cu coeficientul unghiular negativ şi cu ordonata la origine pozitivă; iar (20.76) devine:

$$\frac{dy}{ydt} = A(b-y) \qquad (20.78)$$

sau

$$\frac{dy}{y(b-y)} = Adt \qquad (20.79)$$

sau

$$\frac{bdy}{y(b-y)} = Bdt \qquad (20.80)$$

Pe baza relaţiei (20.80) rezultă:

$$\frac{b}{y(b-y)} = \frac{b-y+y}{y(b-y)} = \frac{1}{y} + \frac{1}{b-y} \qquad (20.81)$$

şi (20.80) devine:

$$(\frac{1}{y} + \frac{1}{b-y})dy = Bdt \qquad (20.82)$$

iar prin integrare, relaţia (20.82) devine:

$$\ln y - \ln(b-y) = Bt + C \qquad (20.83)$$

sau

$$\frac{y}{b-y} = e^{Bt+C} = e^C . e^{Bt} \qquad (20.84)$$

Notînd $e^C = k$, rezultă $C = \ln k$, iar relaţia (20.84) se scrie:

$$\frac{y}{b-y} = ke^{Bt} \qquad (20.85)$$

sau

$$y = bke^{Bt} - yke^{Bt}$$

$$y(1+ke^{Bt}) = bke^{Bt}$$

de unde

$$y = \frac{bke^{Bt}}{1 + ke^{Bt}} \qquad (20.86)$$

Impărţind numărătorul şi numitorul membrului al doilea al relaţiei (20.85) prin ke^{Bt}, rezultă:

$$y = \frac{b}{\frac{1}{ke^{Bt}} + 1} \qquad (20.87)$$

Notînd $a = \frac{1}{k}$, relaţia (20.86) ia forma binecunoscută a ecuaţiei curbei logistice:

$$y = \frac{b}{1 + ae^{-Bt}} \qquad (20.88)$$

Curba logistică dată de ecuaţia (20.88) are ca asimptotă inferioară dreapta $y = 0$, iar ca asimptotă superioară $y = b$ şi punctul de inflexiune $t = 0$ şi $y = \frac{b}{2}$ (fig. 20.11).

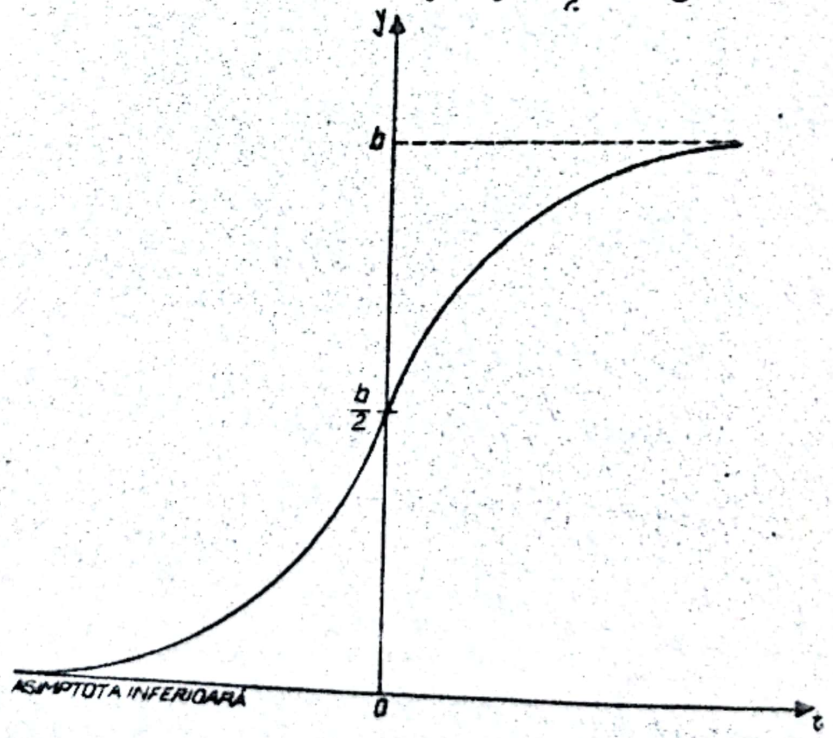


Fig. 20.11. CURBA LOGISTICĂ $y = \frac{b}{1+ae^{-Bt}}$

Inversînd relaţia (20.84) şi logaritmînd, obţinem:

$$\ln \frac{b - y}{y} = - Bt - C \qquad (20.89)$$

Pentru determinarea paramentrilor B, b şi C se poate utiliza metoda punctelor selecţionate. Selectînd trei puncte echidistante în timp $(0,y_1)$, (t,y_2), $(2t,y_3)$ şi punînd curbei (20.89) condiţia de e trece prin cele trei puncte; rezultă sistemul de ecuaţii:

$$\begin{aligned} \ln \frac{b-y_1}{y_1} &= -C \\ \ln \frac{b-y_2}{y_2} &= -Bt-C \\ \ln \frac{b-y_3}{y_3} &= -2Bt-C \end{aligned} \qquad (20.90)$$

Din prima ecuaţie a sistemului (20.90) rezultă valoarea parametrului C :

$$-C = \ln \frac{b-y_1}{y_1} = \frac{1}{0,43429} \log \frac{b-y_1}{y_1} \qquad (20.91)$$

Inmulţind cu -2 a doua ecuaţie a sistemului (20.89) şi apoi adunînd cele trei ecuaţii, se elimină B şi C şi rezultă:

$$\ln \frac{b-y_1}{y_1} - 2 \ln \frac{b-y_2}{y_2} + \ln \frac{b-y_3}{y_3} = 0 \quad , \quad \text{sau}$$

$$\frac{b-y_1}{y_1} \cdot \frac{b-y_3}{y_3} = \left(\frac{b-y_2}{y_2}\right)^2 \quad \text{şi, efectuînd calculele:}$$

$$b^2(y_2^2-y_1y_3)-b(y_1y_2^2+ y_2^2y_3 - 2y_1y_2y_3) = 0$$

Deoarece soluţia $b = 0$ este exclusă, rezultă:

$$b = \frac{2y_1y_2y_3 - y_2^2(y_1 + y_3)}{y_1y_3 - y_2^2} \qquad (20.92)$$

Valoarea parametrului C se obţine substituind pe b

în formula (20.91). In sfîrşit, scăzînd primele două ecuaţii ale sistemului (20.90) obţinem valoarea parametrului B :

$$\ln \frac{b-y_1}{y_1} - \ln \frac{b-y_2}{y_2} = Bt \quad \text{sau} \quad \ln \frac{b-y_1}{y_1} \cdot \frac{y_2}{b-y_2} = Bt \text{ , adică}$$

$$B = \frac{1}{0,43429} \log \frac{(b-y_1)y_2}{y_1(b-y_2)} \qquad (20.93)$$

Un exemplu interesant de aplicare a curbei logistice este prezentat de portughezul Monterio Alves (1965) cu privire la evoluţia preţurilor vinului roşu în Portugalia [1] în perioada 1935 - 1961.

Valorile celor trei parametri ai curbei logistice, determinate pe baza metodei punctelor selecţionate, depind numai de coordonatele celor trei puncte selecţionate şi fac abstracţie de toate celelalte puncte ale evoluţiei reale a fenomenului studiat. Acest fapt constituie principala deficienţă a metodei punctelor selecţionate.

Revenind la expresia generală a curbei logistice, se observă că, inversînd relaţia (20.87), o putem aduce la următoarea formă:

$$\frac{1}{y} = \frac{1}{b} + \frac{e^c}{b} e^{Bt} \qquad (20.94)$$

Notînd acum $\frac{e^c}{b} = \alpha$ şi $e^B = \beta$, relaţia de mai sus ia forma:

$$\frac{1}{y} = \frac{1}{b} + \alpha\beta^t \qquad (20.95)$$

Relaţia (20.95) evidenţiază faptul că diferenţele de ordinul întîi ale valorilor reciproce ale tendinţei variază cu o raţie constantă β . Valoarea lui α este pozitivă, iar β este de asemenea pozitiv şi subunitar.

Ca şi în cazul curbei lui Gompertz, dacă se reprezintă grafic pe o scară semilogaritmică, curba logistică este concavă, ceea ce înseamnă că rata de creştere devine din ce în ce mai mică.

---

1 Alves Monterio, "O Mercado do Vinho em Portugal", Fundaçao Calouste Gulbenkian, Lisboa, 1965.

Pornind de la ecuaţia (20.95), curba logistică se poate liniariza. Astfel, prin logaritmare obţinem:

$$\log\left(\frac{1}{y}-\frac{1}{b}\right)=\log\alpha+t\log\beta \qquad (20.96)$$

sau

$$\log\frac{b-y}{b}=\log\alpha+t\log\beta \qquad (20.97)$$

Această ultimă relaţie evidenţiază faptul că variaţia (creşterea) relativă a funcţiei este o funcţie liniară de timp. Această constatare reprezintă probabil proprietatea cea mai importantă a curbei logistice.

O altă formă a curbei logistice, utilizată pentru determinarea tendinţelor generale de dezvoltare ale fenomenelor economice, este dată de următoarea ecuaţie:

$$\frac{1}{y_{t_i}}=a+bc^{t_i} \qquad (20.98)$$

Din această relaţie se observă asemănarea dintre ecuaţia curbei logistice şi cea a unei curbe exponenţiale modificate (20.50) cu deosebirea că pentru variabila independentă în cazul unei curbe logistice se consideră în locul valorilor ei reale, mărimile inverse ale acestora.

Parametrii a,b şi c ai curbei logistice (20.98) se pot determina pe baza metodei mediilor selecţionate cu ajutorul formulelor:

$$c=\sqrt[n]{\frac{\sum_{2n}^{3n-1}\frac{1}{y_i}-\sum_{n}^{2n-1}\frac{1}{y_i}}{\sum_{n}^{2n-1}\frac{1}{y_i}-\sum_{o}^{n-1}\frac{1}{y_i}}} \qquad (20.99)$$

$$b=\frac{\sum_{n}^{2n-1}\frac{1}{y_i}-\sum_{o}^{n-1}\frac{1}{y_i}}{\sum_{n}^{2n-1}c^{t_i}-\sum_{o}^{n-1}c^{t_i}} \qquad (20.100)$$

$$a=\frac{1}{n}\sum_{o}^{n-1}\frac{1}{y_i}-\frac{b}{n}\sum_{o}^{n-1}c^{t_i} \qquad (20.101)$$

Sistematizat, calculele de ajustare a unei serii dinamice după metoda mediilor selecţionate se prezintă în tabelul 20.15.

Tabelul 20.15. Ajustarea seriilor dinamice după metoda mediilor selecţionate

Anii (t_i)	y_i	$\frac{1}{y_i}$	Mediile selecţionate	$\frac{1}{y_t} = a + bc^{t_i}$	y_t
1	2	3	4	5	6
t_o	y_o	$1/y_o$			
t_1	y_1	$1/y_1$			
.			$\frac{1}{n}\sum_{o}^{n-1}\frac{1}{y_i}$		
.					
t_{n-1}	y_{n-1}	$1/y_{n-1}$			
Total 1		$\sum_{o}^{n-1}\frac{1}{y_i}$			
t_n	y_n	$1/y_n$			
t_{n+1}	y_{n+1}	$1/y_{n+1}$			
.			$\frac{1}{n}\sum_{n}^{2n-1}\frac{1}{y_i}$		
.					
t_{2n-1}	y_{2n-1}	$1/y_{2n-1}$			
Total 2		$\sum_{n}^{2n-1}\frac{1}{y_i}$			
t_{2n}	y_{2n}	$1/y_{2n}$			
t_{2n+1}	y_{2n+1}	$1/y_{2n+1}$			
.			$\frac{1}{n}\sum_{2n}^{3n-1}\frac{1}{y_i}$		
.					
t_{3n-1}	y_{3n-1}	$1/y_{3n-1}$			
Total 3		$\sum_{2n}^{3n-1}\frac{1}{y_i}$			

Să considerăm, pentru exemplificare, seria din tabelul 20.13., referitoare la producţia de oţel a României în perioada 1950 - 1976, a cărei expresie grafică se prezintă în fig. 20.12.

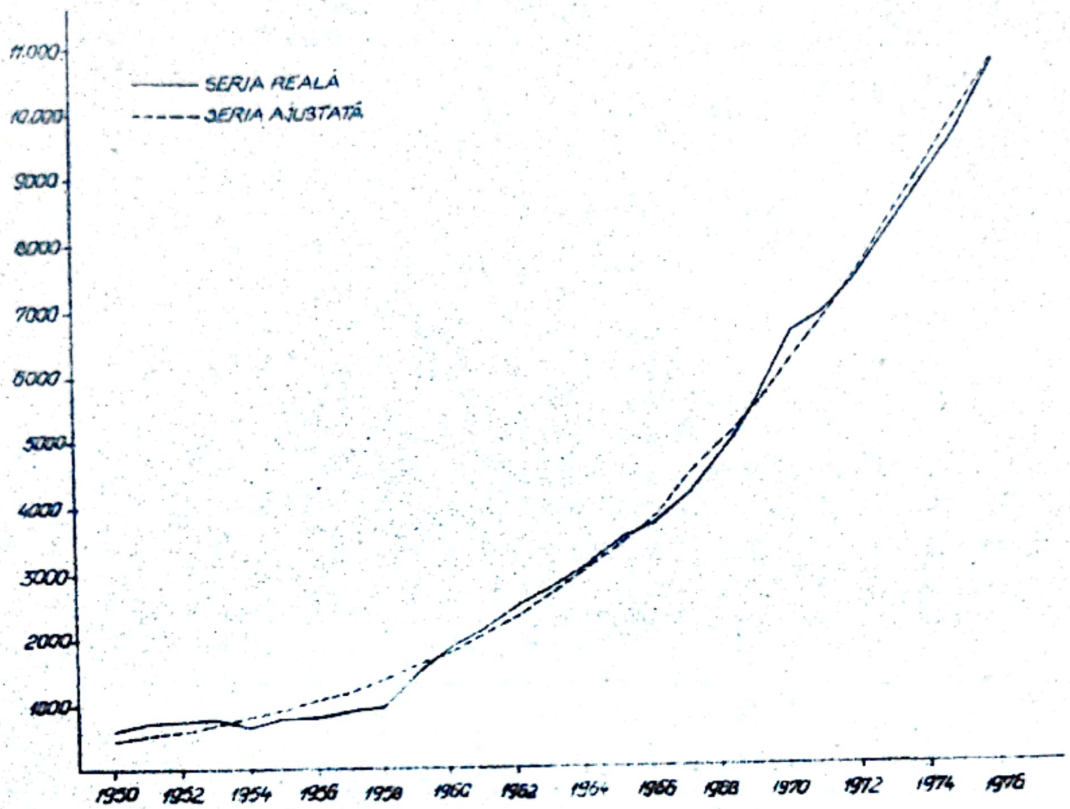


Fig.20.12. PRODUCȚIA DE OȚEL A ROMÂNIEI ÎN PERIOADA 1950 - 1976

Intrucît seria este formată din 27 termeni (număr divizibil prin 3), o vom împărți în trei părți egale, formate din cîte n = 9 termeni, respectiv corespunzătoare perioadelor 1950-1958, 1959-1967 și 1968-1976.

Sumele celor trei grupe de termeni înmulțiți cu 1000 sînt:

$$\sum_{0}^{n-1} \frac{1000}{y_1} = 12,592$$

$$\sum_{n}^{2n-1} \frac{1000}{y_1} = 3,644$$

$$\sum_{2n}^{3n-1} \frac{1000}{y_1} = 1,260$$

Substituind aceste valori în formulele (20.99), (20.100) și (20.101), rezultă:

$$c = \sqrt[9]{\frac{1,260 - 3,644}{3,644 - 12,592}} = \sqrt[9]{0,0266} = 0,863$$

$$b = \frac{3,644 - 12,592}{1,4302 - 5,3476} = 2,272$$

$$a = \frac{1}{9} \cdot 12,592 - \frac{2,272}{9} \cdot 5,3476 = 0,0044$$

Substituind valorile celor trei parametri în ecuația (20.98), rezultă valorile teoretice ale variabilei dependente, care s-au scris în col. 8 a tabelului 20.16.

Tabelul 20.16. Determinarea curbei logistice de evoluție a producției de oțel a României în perioada 1950-1976

Anul	y_i	$\frac{1000}{y_i}$	t_i	c^{t_i}	bc^{t_i}	$a+bc^{t_i}$	y_{t_i}	$y_i - y_{t_i}$
1	2	3	4	5	6	7	8	9
1950	555	1,8018	0	1,0000	2,2725	2,3162	432	123
1951	644	1,5528	1	0,8633	1,9619	2,0057	499	145
1952	698	1,4327	2	0,7454	1,6938	1,7376	576	122
1953	717	1,3947	3	0,6435	1,4623	1,5061	664	53
1954	628	1,5924	4	0,5556	1,2625	1,3062	766	-138
1955	766	1,3055	5	0,4796	1,0899	1,1337	882	-116
1956	779	1,2837	6	0,4141	0,9410	0,9848	1015	-236
1957	864	1,1574	7	0,3575	0,8124	0,8562	1168	-304
1958	934	1,0707	8	0,3086	0,7014	0,7451	1342	-408
1959	1420	0,7042	9	0,2665	0,6055	0,6493	1540	-120
1960	1806	0,5537	10	0,2300	0,5228	0,5665	1765	41
1961	2126	0,4704	11	0,1986	0,4513	0,4951	2020	106
1962	2451	0,4080	12	0,1715	0,3897	0,4334	2307	144
1963	2704	0,3698	13	0,1480	0,3364	0,3802	2630	74
1964	3039	0,3291	14	0,1278	0,2904	0,3342	2992	47
1965	3426	0,2919	15	0,1103	0,2507	0,2945	3395	31
1966	3670	0,2725	16	0,0953	0,2165	0,2602	3843	-173
1967	4088	0,2446	17	0,0822	0,1869	0,2307	4335	-247
1968	4751	0,2105	18	0,0710	0,1614	0,2051	4875	-124
1969	5540	0,1805	19	0,0613	0,1393	0,1831	5462	78
1970	6517	0,1534	20	0,0529	0,1203	0,1640	6096	421
1971	6803	0,1470	21	0,0457	0,1038	0,1476	6775	28
1972	7401	0,1351	22	0,0394	0,0896	0,1334	7496	-95
1973	8161	0,1225	23	0,0341	0,0774	0,1212	8254	-93
1974	8848	0,1130	24	0,0294	0,0668	0,1106	9043	-195
1975	9549	0,1047	25	0,0254	0,0577	0,1015	9857	-308
1976	10733	0,0932	26	0,0219	0,0498	0,0936	10687	46

Scanned with OKEN Scanner

Deci, ecuaţia logistică pentrn creşterea producţiei de oţel a României în perioada 1950- 1976 are forma:

$$\frac{1}{y_t} = 0,044 + 2,272 \cdot 0,863^{t_i}$$

$t_0 = 0$; $y_{1950} = 1/(0,044+2,272.0,863^{0}) = 432$ mii tone
$t_1 = 1$; $y_{1951} = 1/(0,044+2,272.0,863^{1}) = 499$ mii tone
. . . . . .
$t_{26} = 26$; $y_{1976} = 1/(0,044+2,272.0,863^{26}) = 10687$ mii tone

In ipoteza că şi în viitorul apropiat producţia de oţel va fi determinată de influenţa aceloraşi factori care au acţionat în perioada 1950 - 1976, prognoza producţiei de oţel se poate face prin extrapolare, astfel:

$t = 27$; $y_{1977} = 1/(0,044+2,272.0,863^{27}) = 11500$ mii tone
$t = 28$; $y_{1978} = 1/(0,044+2,272.0,863^{28}) = 12333$ mii tone
$t = 29$; $y_{1979} = 1/(0,044+2,272.0,863^{29}) = 13156$ mii tone
$t = 30$; $y_{1980} = 1/(0,044+2,272.0,863^{30}) = 13960$ mii tone etc.

20.2.6.8. Curba exponenţială de gradul doi

Curba exponenţială de gradul doi este descrisă de ecuaţia:

$$y = A.B^{t}.C^{t^2} \qquad (20.102)$$

Logaritmînd ecuaţia (20.102) rezultă:

$$\log y = \log A + t \log B + t^2 \log C \qquad (20.103)$$

Dacă notăm: $a = \log A$, $b = \log B$ şi $C = \log C$, ecuaţia (20.103) devine:

$$\log y = a + bt + ct^2 \qquad (20.104)$$

care se mai numeşte şi parabola logaritmică.

Determinarea parametrilor ecuaţiei exponenţiale de gradul doi (20.103) este similară cu determinarea parametrilor parabolei de gradul doi, cu excepţia faptului că se lucrează cu

logaritmii valorilor lui y în locul valorilor reale ale acestei variabile.

Aplicînd metoda celor mai mici pătrate, rezultă următorul sistem de ecuaţii normale:

$$\begin{aligned} na + b\sum t_i + c\sum t_i^2 &= \sum \log y_i \\ a\sum t_i + b\sum t_i^2 + c\sum t_i^3 &= \sum t_i \log y_i \\ a\sum t_i^2 + b\sum t_i^3 + c\sum t_i^4 &= \sum t_i^2 \log y_i \end{aligned} \qquad (20.105)$$

Pentru a se simplifica calculul parametrilor a, b şi c, substituim sumele puterilor impare ale lui t_i cu zero în (20.105):

$$\begin{aligned} na + c\sum t_i^2 &= \sum \log y_i \\ b\sum t_i^2 &= \sum t_i \log y_i \\ a\sum t_i^2 + c\sum t_i^2 \log y_i & \end{aligned} \qquad (20.106)$$

Din ecuaţia a doua a sistemului (20.106) obţinem:

$$b = \frac{\sum t_i \log y_i}{\sum t_i^2} \qquad (20.107)$$

iar din prima şi a treia ecuaţie a aceluiaşi sistem, rezultă:

$$a = \frac{\sum t_i^4 \sum \log y_i - \sum t_i^2 \sum t_i^2 \log y_i}{n\sum t_i^4 - (\sum t_i^2)^2} \qquad (20.108)$$

$$c = \frac{n\sum t_i^2 \log y_i - \sum t_i^2 \sum \log y_i}{n\sum t_i^4 - (\sum t_i^2)^2} \qquad (20.109)$$

x

x x

Determinarea tendinţei centrale de dezvoltare a fenomenelor este importantă atît pentru descrierea evoluţiei de lungă durată a acestora, cît şi pentru estimarea influenţei factorilor care determină o anumită evoluţie (cum ar fi factorii cu acţiune sezonieră ciclică şi factorii întîmplători).

Ajustînd o serie dinamică nu trebuie să pierdem din vedere faptul că, pentru a satisface cerinţele legii numerelor mari, trebuie să luăm în calcul o perioadă suficient de lungă din trecut. Tendinţa centrală determinată pe baza unui număr redus de termeni este lipsită de semnificaţie şi neindicată pentru prognoză (extrapolare). Luînd în calcul un număr mai mare de termeni, este posibil să se determine adevăratele cauze ale evoluţiei, altele decît variabila independentă timp.

Experienţa statistică a arătat că <u>orizontul prognozei</u> obţinute pe baza tendinţei trebuie să fie foarte mic în comparaţie cu durata luată în calculul tendinţei. Astfel, de exemplu, dacă tendinţa s-a determinat pentru o perioadă de 50 de ani, atunci orizontul prognozei trebuie să fie mai mic de un deceniu.

Calitatea ajustării se poate aprecia pe baza comparaţiei între abaterea medie pătratică a valorilor teoretice y_t faţă de valorile empirice y_i şi abaterea medie pătratică a valorilor empirice faţă de valoarea medie $\bar{y}$. Dacă perioada de timp luată în calcul este suficient de mare, atunci σ_{yy_t} este mult mai mică decît σ_y.

I N D I C I

21.1.Noţiunea şi impertanţa indicilor

Studiul variaţiei fenomenelor, în general, şi a variaţiei lor în timp, în special, i-a preocupat pe oamenii de ştiinţă din totdeauna dar primele începuturi ale tratării lor ştiinţifice au apărut de abia spre mijlocul secolului al XVIII-lea, odată cu încercarea lui Dutot, din anul 1738,de a compara sumele preţurilor din două perioade, fapt pentru care el a fost considerat uneori, ca fiind fondatorul numerelor indice.

Un sfert de secol mai tîrziu, italianul G.R.Carli[x)] a construit un indice al preţurilor sub formă de medie aritmetică simplă a indicilor individuali, pentru a măsura scăderea puterii de cumpărare a banilor ca urmare a importului de argint în Europa din America. Carli a considerat doar preţurile la trei produse principale (grîu, vin şi ulei) în anii 1500 şi 1750 şi a calculat variaţia lor relativă în decursul celor 250 de ani consideraţi. Apoi a determinat media aritmetică simplă a celor trei mărimi relative, obţinînd în felul acesta un indice al preţurilor care, după cum se ştie, se apropie de mărimea inversă a puterii de cumpărare a banilor. De aceea, Carli este aproape unanim recunoscut în istoria statisticii drept inventatorul numerelor indice.

Contribuţii mai importante în teoria şi practica indicilor au adus în perioada modernă Herman Paasche, Etienne Laspeyres, Ladislau von Bortkiewicz,W.S.Jevons şi F.Y. Edgeworth (1845-1926),A.L.Bowley (1859-1957) în Anglia, C.M. Walsh, A.C.Pigou, Irving Fisher, Bruce Mudgett, W.C.Mitchell, A.A.Young, C.Gini, F.Bandettini, L.Kazineţ, S.G.Strumilin, L.Tövissi şi alţii.

---

x)Carli G.R.,"Del valore e della proporzione de metalli monetati con i generi in Italia prima delle Scoperte dell Indice col confronto del valore e della proporzione de tempi nostri", 1764, publicat de Custodi în: Scrittori italiani de economia politica, parte moderna, vol.XIII, pp. 297- 366.

Indicii reprezintă o categorie specială a indicatorilor statistici cu ajutorul cărora se măsoară variaţia medie relativă a fenomenelor sociale şi economice în timp şi în spaţiu sau în raport cu un nivel planificat.

Prin definiţie orice indice se obţine prin raportarea între ele a două niveluri absolute ale aceluiaşi fenomen sau grup de fenomene. Unul dintre niveluri reprezintă volumul unui fenomen din perioada care se raportează, iar celălalt reprezintă volumul fenomenului din perioada considerată ca bază de raportare.

Alegerea bazei de raportare se face după aceleaşi principii ca şi în cazul ritmului mediu. Nivelul ales ca bază de comparaţie trebuie să reprezinte un nivel normal din dezvoltarea fenomenului cercetat, un nivel care să aibă o semnificaţie deosebită (să reprezinte o anumită etapă sau treaptă de dezvoltare) în aşa fel încît să justifice alegerea lui ca etalon în analiza statistică.

La construirea unui indice trebuie să se ţină seama de următoarele reguli:

1)Orice indice este rezultatul comparării a două situaţii sau a unei comparaţii binare cum se mai numeşte uneori în literatura de specialitate. Prin generalizare se poate ajunge la compararea în serie. Cele două situaţii din comparaţie binară pot fi două perioade de timp, două localităţi sau zone diferite.

2)Informaţiile iniţiale necesare comparaţiei binare cu privire la toate bunurile economice luate în studiu trebuie să fie complete şi precise. De exemplu, dacă se calculează un indice al preţurilor pentru o serie de produse este necesar să se cunoască cantităţile şi preţurile fiecărui produs în cele două momente sau perioade de comparat. Prin înmulţirea cantităţilor q_i cu preţurile p_i, pentru fiecare produs, se obţin valorile $v_i = q_i p_i$, iar prin împărţirea sumei valorilor tuturor produselor $\sum_i v_i = \sum_i q_i p_i$ din cele două perioade se obţine un indice al cărui valoare este, de regulă, diferită de unitate.

Deoarece indicii se obţin prin raportarea nivelului atins de unul sau mai multe fenomene în două momente sau perioade diferite, ei sînt indicatori fără dimensiune, adică independenţi de unitatea de măsură a fenomenului luat în calcul.

21.2.Clasificarea indicilor

In teoria şi practica statistică se utilizează numeroşi indici care diferă între ei din multiple puncte de vedere. De aceea, pentru formarea unei imagini coerente, de ansamblu asupra lor, este necesar să-i clasificăm după criterii diferite cum ar fi: sfera de cuprindere, natura caracteristicii, modul de alegere a bazei de comparaţie, procedeul de calcul etc.

Din punct de vedere al sferei de cuprindere indicii se împart în două categorii:

-indici individuali, simpli sau elementari care măsoară variaţia relativă a unor fenomene simple, singulare, în timp sau în spaţiu;

-indici de grup, compuşi sau sintetici care măsoară variaţie medie relativă a unor fenomene colective în raport cu o anumită bază de comparaţie.

După natura caracteristicii studiate se disting trei categorii de indici:

-indici ai variaţiei în timp sau indici ai dinamicii;
-indici ai variaţiei în spaţiu sau indici teritoriali;
-indici ai variaţiei în raport cu planul sau indici de plan.

In funcţie de modul de alegere a bazei de raportare atît indicii individuali cît şi cei de grup se împart în:

-indici cu bază fixă;
-indici cu bază mobilă;

Din punct de vedere al procedeului de calcul indicii de grup se împartr în:

-indici agregaţi;

-indici calculaţi ca medii aritmetice ponderate a indicilor individuali;

-indici calculaţi ca medii armonice ponderate a indicilor individuali;

-indici calculaţi prin raportarea a două medii aritmetice ponderate;

-indici calculaţi sub formă de medie geometrică etc.

21.3.Indicii dinamici

Intrucît indicii dinamicii reprezintă categoria cea mai importantă şi cea mai dezvoltată din cadrul tuturor indicilor vom începe prezentarea conţinutului metodologic al indicilor cu această categorie.

Aşa după cum reiese şi din denumirea lor indicii dinamicii măsoară în variaţia medie relativă a fenomenelor în timp. Intervalul de timp cuprins între cele două momente sau perioade de comparat trebuie astfel ales încît în cadrul lui fenomenul cercetat să nu fi suferit modificări calitative importante în aşa fel încît fenomenul raportat să fie într-adevăr comparabil cu nivelul lui atins în perioada de bază.

21.3.1.Indici simpli sau elementari

Să considerăm cantităţile şi preţurile a "n" produse într-o perioadă de k + 1 ani(tabelul 21.1).

Tabelul 21.1

Anii	Cantităţile produsului					Preţurile produsului				
	1	2	3	...j	...n	1	2	3	...j	...n
o	q_{o1}	q_{o2}	q_{o3}	q_{oj}	q_{on}	p_{o1}	p_{o2}	p_{o3}	p_{oj}	p_{on}
1	q_{11}	q_{12}	q_{13}	q_{1j}	q_{1n}	p_{11}	p_{12}	p_{13}	p_{1j}	p_{1n}
2	q_{21}	q_{22}	q_{23}	q_{2j}	q_{2n}	p_{21}	p_{22}	p_{23}	p_{2j}	p_{2n}
3	q_{31}	q_{32}	q_{33}	q_{3j}	q_{3n}	p_{31}	p_{32}	p_{33}	p_{3j}	p_{3n}
...	...	...	...	...	...	...	...	...	...	...
i	q_{i1}	q_{i2}	q_{i3}	q_{ij}	q_{in}	p_{i1}	p_{i2}	p_{i3}	p_{ij}	p_{in}
...	...	...	...	...	...	...	...	...	...	...
k	q_{k1}	q_{k2}	q_{k3}	q_{kj}	q_{kn}	p_{k1}	p_{k2}	p_{k3}	p_{kj}	p_{kn}

Din tabelul 21.1 rezultă că q_{ij} şi p_{ij} reprezintă respectiv cantitatea şi preţul produsului j în anul i.

Valoarea produsului j în anul i va fi egală cu produsul $q_{ij}\ p_{ij}$, adică se înmulţeşte linia i din coloana a doua a tabelului cu linia i din coloana a treia din acelaşi tabel.

Pentru cunoaşterea variaţiei în timp a volumului fizic al fiecărui produs sau a dinamicii volumului fizic se calculează indicii individuali ai volumului fizic. In funcţie de modul de alegere a bazei de comparaţie indicii individuali pot fi de două feluri:

-cu bază fixă - notat cu $i^{q}_{i/o}$;

-cu bază mobilă (sau în lanţ)-notat cu $i^{q}_{i/i-1}$.

Indicii individuali cu bază fixă ai volumului fizic se calculează prin raportarea volumului fizic al produsului j din anul i la volumul fizic al aceluiaşi produs din anul de bază. Dacă se consideră ca an de bază anul "o" atunci se obţin următorii indici individuali ai volumului fizic şi ai preţurilor în perioada "i" faţă de "o":

- pentru produsul j în perioada "1" faţă de "o" :

$$i^{q_j}_{1/o} = \frac{q_{1j}}{q_{oj}}\cdot 100 \text{ şi } i^{p_j}_{1/o} = \frac{p_{1j}}{p_{oj}}\cdot 100 \qquad j=1,2,\ldots,n$$

-pentru produsul j în perioada "2" faţă de "o":

$$i^{q_j}_{2/o} = \frac{q_{2j}}{q_{oj}}\cdot 100 \text{ şi } i^{p_j}_{2/o} = \frac{p_{2j}}{p_{oj}}\cdot 100 \qquad j=1,2,\ldots,n$$

. . . . . .

-pentru produsul j în perioada "i" faţă de "o":

$$i^{q_j}_{i/o} = \frac{q_{ij}}{q_{oj}}\cdot 100 \text{ şi } i^{p_j}_{i/o} = \frac{p_{ij}}{p_{oj}}\cdot 100 \qquad \begin{matrix} i=1,2,\ldots,k \\ j=1,2,\ldots,n \end{matrix} \qquad (21.1)$$

. . . . . .

-pentru produsul j în perioada "k" faţă de "o":

$$i^{q_j}_{k/o} = \frac{q_{kj}}{q_{oj}}\cdot 100 \text{ şi } i^{p_j}_{k/o} = \frac{p_{kj}}{p_{oj}}\cdot 100 \qquad j=1,2,\ldots,n$$

Indicii individuali cu bază mobilă se obţin prin schimbarea bazei de la o perioadă la alta, astfel:

Scanned with OKEN Scanner

-pentru produsul j în perioada "i" față de "o":

$$i^{q_j}_{i/o} = \frac{q_{ij}}{q_{oj}}\cdot 100 \text{ și } i^{p_j}_{i/o} = \frac{p_{ij}}{p_{oj}}\cdot 100 \qquad j=1,2,\ldots,n$$

-pentru produsul j în perioada "2" față de "1":

$$i^{q_j}_{2/1} = \frac{q_{2j}}{q_{1j}}\cdot 100 \text{ și } i^{p_j}_{2/1} = \frac{p_{2j}}{p_{1j}}\cdot 100 \qquad j=1,2,\ldots,n$$

. . . . . .

-pentru produsul j în perioada "i" față de "i-1":

$$i^{q_j}_{i/i-1} = \frac{q_{ij}}{q_{i-1,j}}\cdot 100 \text{ și } i^{p_j}_{i/i-1} = \frac{p_{ij}}{p_{i-1,j}}\cdot 100 \qquad \begin{matrix} i=1,2,..,k \\ j=1,2,..,n \end{matrix} \qquad (21.2)$$

. . . . . .

-pentru produsul j în perioada "k" față de "k-1":

$$i^{q_j}_{k/k-1} = \frac{q_{kj}}{q_{k-1,j}}\cdot 100 \text{ și } i^{p_j}_{k/k-1} = \frac{p_{kj}}{p_{k-1,j}}\cdot 100 \qquad j=1,2,\ldots,n$$

Între indicii individuali cu bază fixă și indicii individuali cu bază mobilă există o relație, bine determinată, și anume, produsul indicilor cu bază mobilă este egal cu indicele cu bază fixă al întregii perioade:

$$\prod_{i=1}^{k} i^{q_j}_{i/i-1} = i^{q_j}_{k/o} \text{ și } \prod_{i=1}^{k} i^{p_j}_{i/i-1} = i^{p_j}_{k/o} \qquad j=1,2,\ldots,n \qquad (21.3)$$

Din relația (21.3)rezultă că raportul dintre doi indici cu bază fixă este egal cu indicele cu bază mobilă corespunzător perioadelor de raportare:

$$i^{q_j}_{i/o} : i^{q_j}_{i-1/o} = i^{q_j}_{i/i-1} \text{ și } i^{p_j}_{i/o} : i^{p_j}_{i-1/o} = i^{p_j}_{i/i-1} \qquad j=1,2,..,n \qquad (21.4)$$

Relațiile (21.3) și (21.4) definesc tocmai una dintre proprietățile cele mai importante ale indicilor individuali

şi anume cea de <u>circularitate</u>. Pe baza proprietăţii de circularitate se poate face trecerea de la indicii cu bază fixă la indicii cu bază mobilă şi invers.

O altă proprietate a indicilor individuali este cea de <u>reversibilitate</u>, care se poate scrie astfel:

$$i_{i/o} = \frac{1}{i_{o/i}} \qquad (21.5)$$

adică, în cazul volumului fizic:

$$i^{q_j}_{i/o} = \frac{1}{i^{q_j}_{o/i}} = \frac{1}{\frac{q^j_o}{q^j_i}} = \frac{q^j_i}{q^j_o}$$

In mod analog se vor obţine indicii individuali ai valorii prin compararea valorii unui produs în două perioade diferite. Astfel:

-indicii cu bază fixă a valorii produsului j:

$$i^{v_j}_{i/o} = \frac{v_{ij}}{v_{oj}} \cdot 100 = \frac{q_{ij}\, p_{ij}}{q_{oj}\, p_{oj}} \cdot 100 \qquad \begin{matrix} i=0,1,2,\ldots,k \\ j=1,2,3,\ldots,n \end{matrix} \qquad (21.6)$$

deci

$$i^{v_j}_{i/o} = \frac{q_{ij}}{q_{oj}} \cdot \frac{p_{ij}}{p_{oj}} = i^{q_j}_{i/o} \cdot i^{p_j}_{i/o} \qquad (21.7)$$

-indicii cu bază mobilă ai valorii produsului j:

$$i^{v_j}_{i/i-1} = \frac{v_{ij}}{v_{i-1,j}} \cdot 100 = \frac{q_{ij}\, p_{ij}}{q_{i-1,j} p_{i-1,j}} \cdot 100 \qquad \begin{matrix} i=0,1,2,\ldots,k \\ j=1,2,\ldots\ldots,n \end{matrix} \qquad (21.8)$$

$$i^{v_j}_{i/i-1} = \frac{q_{ij}}{q_{i-1,j-1}} \cdot \frac{p_{ij}}{p_{i-1,j-1}} = i^{q_j}_{i/i-1} \cdot i^{p_j}_{i/i-1} \qquad (21.9)$$

In mărimi absolute influenţa ambilor factori de variaţie asupra valorii fiecărui produs, calculată pe baza indicelui individual valoric, este egală cu suma influenţelor celor doi factori:

$$\Delta^{v_j}_{i/o} = \Delta^{q_j}_{i/o} + \Delta^{p_j}_{i/o} \text{ şi } \Delta^{v_j}_{i/i-1} = \Delta^{q_j}_{i/i-1} - \Delta^{p_j}_{i/i-1} \qquad (21.10)$$

adică

$$q_{ij}\,p_{ij} - q_{oj}\,p_{oj} = (q_{ij} - q_{oj}) + (p_{ij} - p_{oj})$$

şi respectiv

$$q_{ij}\,p_{ij} - q_{i-1,j}\,p_{i-1,j} = (q_{ij} - q_{i-1,j}) + (p_{ij} - p_{i-1,j})$$

Din relaţiile (21.7) şi (21.9) se observă că, indiferent de baza de comparaţie aleasă, un indice individual al valorii unui produs este egal cu produsul dintre indicele individual al volumului fizic şi indicele individual al preţului corespunzător.

Valoarea oricărui indice individual poate să fie mai mică, egală sau mai mare decît unitatea sau, decît 100 atunci cînd indicele se exprimă în procente. In general un indice arată de cîte ori a crescut sau a scăzut nivelul unui anumit fenomen în perioada de comparat faţă de nivelul aceluiaşi fenomen din perioada considerată de bază, sau din aceeaşi perioadă de timp dar situat în spaţii diferite. Dacă indicele este mai mic decît unitatea înseamnă că nivelul unui fenomen a scăzut în perioada comparată faţă de perioada de bază; dacă valoarea indicelui este egală cu unitatea înseamnă că fenomenul considerat nu a suferit nici o modificare (în sens de creştere sau scădere) în perioada cercetată; iar dacă valoarea unui indice este supraunitară această arată că fenomenul studiat a crescut în perioada studiată.

Atunci cînd indicii se exprimă în procente, perioada

de bază este echivalată cu 100%, iar valoarea indicelui arată cu cît a crescut sau a scăzut, în procente, nivelul unui anumit fenomen în perioada de comparat față de perioada aleasă ca bază de comparație.

21.3.2.Indici de grup sau sintetici

Cunoașterea variației în timp sau în spațiu a fenomenelor individuale este importantă dar nu este suficientă pentru formarea unei imagini de ansamblu, generale, asupra variației fenomenelor de masă. Dinamica și variația în spațiu a fenomenelor colective se măsoară în statistică cu ajutorul <u>indicilor de grup</u> sau <u>sintetici</u>.

Indicii de grup se notează cu litere majuscule și pot fi cu bază fixă și cu bază mobilă.

21.3.2.1. Indicii agregați

Pentru construcția indicilor agregați să considerăm din nou elementele din tabelul 21.1 și, pe baza-lor, să calculăm volumul valoric al celor n produse în perioada o-k. Rezultatele se prezintă în tabelul 21.2.

Tabelul 21.2

Anii \ Produsul	1	2	....	j	....	n	Total
o	$q_{o1}p_{o1}$	$q_{o2}p_{o2}$	....	$q_{oj}p_{oj}$	....	$q_{on}p_{on}$	$\sum_{o}^{n} q_{oj}p_{oj}$
1	$q_{11}p_{11}$	$q_{12}p_{12}$	....	$q_{1j}p_{1j}$	....	$q_{1n}p_{1n}$	$\sum_{o}^{n} q_{1j}p_{1j}$
2	$q_{21}p_{21}$	$q_{22}p_{22}$	....	$q_{2j}p_{2j}$	....	$q_{2n}p_{2n}$	$\sum_{o}^{n} q_{2j}p_{2j}$
⋮	⋮	⋮		⋮		⋮	⋮
i	$q_{i1}p_{i1}$	$q_{i2}p_{i2}$	....	$q_{ij}p_{ij}$	....	$q_{in}p_{in}$	$\sum_{o}^{n} q_{ij}p_{ij}$
⋮	⋮	⋮		⋮		⋮	⋮
k	$q_{k1}p_{k1}$	$q_{k2}p_{k2}$	....	$q_{kj}p_{kj}$	....	$q_{kn}p_{kn}$	$\sum_{o}^{n} q_{kj}p_{kj}$

Pe ultima coloană a tabelului 21.2 se prezintă volumul valoric total al celor n produse din fiecare an.

In funcţie de modul de alegere a bazei de comparaţie se disting două categorii de indici agregaţi:

-cu bază fixă-notat cu $I_{i/o}$;

-cu bază mobilă-notat cu $I_{i/i-1}$.

Atît indicii cu bază fixă cît şi cei cu bază mobilă, în funcţie de modul de alegere a ponderilor, pot fi cu ponderi constante sau cu ponderi variabile.

Indicii cu bază fixă şi ponderi constante ai volumului fizic şi ai preţurilor se calculează cu ajutorul formulelor:

$$I^{q}_{i/o} = \frac{\sum_j q_{ij}p_{oj}}{\sum_j q_{oj}p_{oj}} \qquad \begin{matrix} i = o,1,2,\ldots, k \\ j = 1,2, \ldots, n \end{matrix} \qquad (21.11)$$

respectiv

$$I^{p}_{i/o} = \frac{\sum_j p_{ij}q_{oj}}{\sum_j p_{oj}q_{oj}} \qquad (21.12)$$

In ambele formule s-au utilizat ca ponderi nivelurile preţurilor şi respectiv ale volumului fizic din anul de bază. In felul acesta sumele de la numitorul formulelor (21.11) şi (21.12) reprezintă valoarea reală a produselor din perioada de bază exprimată în preţurile aceleeaşi perioade, în timp ce $\sum_j q_{ij}p_{oj}$ şi $\sum_j p_{ij}q_{oj}$ reprezintă valori convenţionale ale produselor din perioada de bază şi ale celor din perioada curentă.

Pe baza indicilor de grup în formă agregat se poate determina influenţa variaţiei factorilor luaţi în studiu în perioada i faţă de perioada de bază, în mărimi absolute sub forma diferenţei dintre numărătorul şi numitorul formulei indicelui respectiv.

Astfel, în cazul indicelui de grup al volumului fizic (21.11) rezultă:

$$\Delta^{q}_{i/o} = \sum_{j} q_{ij}p_{oj} - \sum_{j} q_{oj}p_{oj}$$

Acest indicator reprezintă diferenţa dintre suma valorii tuturor celor j produse în anul i exprimate în preţurile perioadei de bază "o" şi valoarea aceloraşi j produse din perioada de bază exprimate în preţurile perioadei de bază.

Din formula (21.12) rezultă:

$$\Delta^{p}_{i/o} = \sum_{j} p_{ij}q_{oj} - \sum_{j} p_{oj}q_{oj}$$

Un indicator important de caracterizare a nivelului de trai al populaţiei îl constituie <u>indicele costului vieţii</u>. Acest indice măsoară variaţia medie relativă a preţurilor mărfurilor şi tarifelor serviciilor care formează consumul.

Indicii cu bază fixă şi cu ponderi variabile ai volumului fizic şi ai preţurilor se calculează astfel:

$$I^{q}_{i/o} = \frac{\sum_{j} q_{ij}\, p_{ij}}{\sum_{j} q_{oj}\, p_{ij}} \qquad \begin{matrix} i=o,1,2,\dots,k \\ j=1,2,\dots,n \end{matrix} \qquad (21.13)$$

respectiv

$$I^{p}_{i/o} = \frac{\sum_{j} p_{ij}\, q_{ij}}{\sum_{j} p_{oj}\, q_{ij}} \qquad (21.14)$$

Indicele care măsoară în mod concomitent variaţia preţurilor şi a volumului fizic al tuturor produselor este <u>indicele agregat al valorii</u>, definit de formula:

$$I^{v}_{i/o} = \frac{\sum_{j} q_{ij}\, p_{ij}}{\sum_{j} q_{oj}\, p_{oj}} \qquad \begin{matrix} i=o,1,2,\dots,k \\ j=o,1,2,\dots,n \end{matrix} \qquad (21.15)$$

şi

$$I^{v}_{i/i-1} = \frac{\sum_j q_{ij}\, p_{ij}}{\sum_j q_{i-1,j}\, p_{i-1,j}} \qquad (21.16)$$

Deoarece indicele agregat al valorii este un indice bifactorial, la fel ca şi indicele individual al valorii se poate descompune în factori. Între indicele valoric, indicele volumului fizic şi indicele preţurilor există relaţia:

$$I^{v}_{i/o} = I^{q}_{i/o} \cdot I^{p}_{i/o} \quad \text{şi} \quad I^{v}_{i/i-1} = I^{q}_{i/i-1} \cdot I^{p}_{i/i-1} \qquad (21.17)$$

adică

$$\frac{\sum_j q_{ij}\, p_{ij}}{\sum_j q_{oj}\, p_{oj}} = \frac{\sum_j q_{ij}\, p_{oj}}{\sum_j q_{oj}\, p_{oj}} \cdot \frac{\sum_j q_{ij}\, p_{ij}}{\sum_j q_{ij}\, p_{oj}} \qquad (21.18)$$

şi respectiv

$$\frac{\sum_j q_{ij}\, p_{ij}}{\sum_j q_{i-1,j}\, p_{i-1,j}} = \frac{\sum_j q_{ij}\, p_{i-1,j}}{\sum_j q_{i-1,j}\, p_{i-1,j}} \cdot \frac{\sum_j q_{ij}\, p_{ij}}{\sum_j q_{ij}\, p_{i-1,j}} \qquad (21.19)$$

În cazul influenţelor factorilor exprimate în mărimi absolute se va verifica relaţia:

$$\Delta^{v}_{1/o} = \Delta^{q}_{1/o} + \Delta^{p}_{1/o} \qquad (21.20)$$

adică

$$\sum_j q_{1j}\, p_{1j} - \sum_j q_{oj}\, p_{oj} = \left(\sum_j q_{1j}\, p_{oj} - \sum_j q_{oj}\, p_{oj}\right) + \left(\sum_j q_{1j}\, p_{1j} - \sum_j q_{1j}\, p_{oj}\right) \qquad (21.21)$$

şi respectiv

$$\sum_j q_{ij}\,p_{ij} - \sum_j q_{i-1,j}\,p_{i-1,j} = \left(\sum_j q_{ij}\,p_{i-1,j} - \sum_j q_{i-1,j}\,p_{i-1,j}\right) + \left(\sum_j q_{ij}\,p_{ij} - \sum_j q_{ij}\,p_{i-1,j}\right) \qquad (21.22)$$

Precizăm faptul că relaţiile (21.17) şi (21.20) se verifică atît în cazul indicelor cu bază fixă (21.18) cît şi în cazul indicelor cu bază mobilă (21.19) dacă unul dintre indicii factorali simpli este ponderat cu nivelul factorului de ponderare din perioada de bază şi celălalt cu nivelul corespunzător din perioada curentă.

La rîndul lor indicii sub formă agregat cu bază mobilă pot fi construiţi cu ponderi constante sau cu ponderi variabile, astfel:

-în cazul utilizării unor ponderi constante:

$$I^{q}_{i/i-1} = \frac{\sum_j q_{ij}\,p_{oj}}{\sum_j q_{i-1,j}\,p_{oj}} \qquad \begin{matrix} i = o,1,2,\ldots,k \\ j = o,1,2,\ldots,n \end{matrix} \qquad (21.23)$$

şi

$$I^{p}_{i/i-1} = \frac{\sum_j p_{ij}\,q_{oj}}{\sum_j p_{i-1,j}\,q_{oj}} \qquad (21.24)$$

-în cazul utilizării unor ponderi variabile:

$$I^{q}_{i/i-1} = \frac{\sum_j q_{ij}\,p_{i-1,j}}{\sum_j q_{i-1,j}\,p_{i-1,j}} \qquad \begin{matrix} i = o,1,2,\ldots,k \\ j = o,1,2,\ldots,n \end{matrix} \qquad (21.25)$$

şi

$$I^{p}_{i/i-1} = \frac{\sum_j p_{ij}\,q_{i-1,j}}{\sum_j p_{i-1,j}\,q_{i-1,j}} \qquad (21.26)$$

-în cazul reflectării concomitente a influenţei ambilor factori:

$$I^{v}_{i/i-1} = \frac{\sum_j q_{ij}p_{ij}}{\sum_j q_{i-1,j}p_{i-1,j}} \qquad (21.27)$$

Statisticianul german Etienne-Laspeyres (1834-1913) a propus ca indicii agregaţi să se construiască cu ajutorul ponderilor din perioada de bază (21.11) şi (21.12).Aceşti indici se utilizează în studierea dinamicii volumului fizic al produselor,încercînd a se elimina influenţa factorului de cemensurare preţ.

Formulele indiciilor agregat (21.13) şi 21.14), în care se utilizează ca ponderi elemente ale perioadei curente,poartă denumirea de indici de tip Paasche, după numele economistului german Herman Paasche (1851-1925) care a propus pentru prima dată acest sistem de ponderare. Indicii de tip Paasche se aplică în practica statistică pentru calculul dinamicii preţurilor întrucît interesează în ce măsură au variat preţurile produselor din perioada curentă.

Din punct de vedere metodologic-ambele formule sînt la fel de satisfăcătoare prezentînd avantaje şi dezavantaje asemănătoare.

Indicele de tip Paasche şi indicele de tip Laspeyres calculaţi pentru acelaşi fenomen vor avea valori diferite în măsura în care ponderile din perioara de bază diferă de cele din perioada curentă.

Diferenţa dintre un indice de tip Paasche şi un indice de tip Laspeyres a fost determinată de către statisticianul german Ladislau von Bortkiewicz (1868-1931). Relaţia lui Bortkiewicz pentru indicii volumului fizic are următoarea expresie:

$$\frac{\sum_j q_{ij}p_{ij}}{\sum_j q_{oj}p_{ij}} : \frac{\sum_j q_{ij}p_{oj}}{\sum_j q_{oj}p_{oj}} = 1 + r_{iq_{i/o}} \cdot i_{p_{i/o}} \cdot v_{iq_{i/o}} \cdot v_{ip_{i/o}} \qquad (21.28)$$

unde

$$r_{i^q_{i/o},\, i^p_{i/o}} = \frac{\sum_j (i^q_{i/o} - I^q_{i/o})(i^p_{i/o} - I^p_{i/o})}{\sigma_{i^q_{i/o}}\ \sigma_{i^p_{i/o}}\ \sum_j q_{oj}p_{oj}}$$

$$\sigma_{i^q_{i/o}} = \frac{\sum_j (i^q_{i/o} - I^q_{i/o})^2\, q_{oj}p_{oj}}{\sum_j q_{oj}p_{oj}}$$

$$\sigma_{i^p_{i/o}} = \frac{\sum_j (i^p_{i/o} - I^p_{i/o})^2 q_{oj}p_{oj}}{\sum_j q_{oj}p_{oj}}$$

Statisticianul englez Francis Ysidor Edgeworth(1845-1926) a propus ca indicele agregat să fie ponderat cu media aritmetică simplă a ponderilor din perioadele de comparat , astfel:

-în cazul volumului fizic:

$$I^q_{i/o} = \frac{\sum_j q_{ij}\frac{p_{oj}+p_{ij}}{2}}{\sum_j q_{oj}\frac{p_{oj}+p_{ij}}{2}} = \frac{\sum_j q_{ij}(p_{oj}+p_{ij})}{\sum_j q_{oj}(p_{oj}+p_{ij})} = \frac{\sum_j q_{ij}p_{oj} + \sum_j q_{ij}p_{ij}}{\sum_j q_{oj}p_{oj} + \sum_j q_{oj}p_{ij}} \qquad (21.29)$$

-în cazul preturilor:

$$I^p_{i/o} = \frac{\sum_j p_{ij}\frac{q_{oj}+q_{ij}}{2}}{\sum_j p_{oj}\frac{q_{oj}+q_{ij}}{2}} = \frac{\sum_j p_{ij}q_{oj} + \sum_j p_{ij}q_{ij}}{\sum_j p_{oj}q_{oj} + \sum_j p_{oj}q_{ij}} \qquad (21.30)$$

Scanned with OKEN Scanner

O altă soluţie în problema ponderării indicilor agregat a fost propusă de statisticianul Drobisch în anul 1871. Drobisch consideră că este necesar să se ia în calculul indicelui agregat atît ponderile din perioada curentă cît şi cele din perioada de bază iar indicele agregat să se calculeze ca o medie aritmetică simplă a indicilor de tip Laspeyres şi a celor de tip Paasche după formula:

-în cazul volumului fizic:

$$I^{q}_{i/o} = \frac{\dfrac{\sum_i q_{ij}p_{oj}}{\sum_i q_{oj}p_{oj}} + \dfrac{\sum_i q_{ij}p_{ij}}{\sum_i q_{oj}p_{ij}}}{2} \qquad (21.31)$$

-în cazul preţurilor:

$$I^{p}_{i/o} = \frac{\dfrac{\sum_i p_{ij}q_{oj}}{\sum_i p_{oj}q_{oj}} + \dfrac{\sum_i p_{ij}q_{ij}}{\sum_i p_{oj}q_{ij}}}{2} \qquad (21.32)$$

Pentru exemplificarea calculului şi analizei indicilor individuali şi a celor de grup sub formă agregat să considerăm datele convenţionale din tabelul 21.3.

Tabelul 21.3.

Produsul	U.M.	Perioada de bază		Perioada curentă	
		Volumul fizic	Preţurile unitare (lei)	Volumul fizic	Preţurile unitare (lei)
A	kg	200	40	300	40
B	m	300	20	300	15
C	buc	500	30	450	35

Variaţia relativă a volumului fizic al fiecărui produs în perioada curentă faţă de perioada de bază este dată de indicii individuali calculaţi după formula (21.1), astfel:

-pentru produsul A:

$$i_{1/0}^{q_j} = \frac{q_1}{q_0} = \frac{300}{200} = 1,5 \quad \text{sau} \quad i_{1/0}^{q_j} \cdot 100 - 100 = 150 - 100 = 50\%$$

iar

$$\Delta_{1/0}^{q_j} = q_1 - q_0 = 300 - 200 = 100 \text{ kg}$$

Deci volumul fizic al produsului A a crescut în perioada curentă față de perioada de bază cu 100 kg sau de 1,5 ori sau, în procente, cu 50%.

-pentru produsul B:

$$i_{1/0}^{q_j} = \frac{300}{300} = 1 \quad \text{sau} \quad i_{1/0}^{q_j} \cdot 100 - 100 = 0$$

iar

$$\Delta_{1/0}^{q_j} = 300 - 300 = 0$$

Aceasta înseamnă că volumul fizic al produsului B a rămas nemodificat în perioada curentă față de perioada de bază.

-pentru produsul C:

$$i_{1/0}^{q_j} = \frac{450}{500} = 0,9 \quad \text{sau} \quad i_{1/0}^{q_j} \cdot 100 - 100 = -10\%$$

iar

$$\Delta_{1/0}^{q_j} = 450 - 500 = -50 \text{ buc.}$$

Volumul fizic al produsului C s-a micșorat în perioada curentă față de perioada de bază cu 50 bucăți sau de 0,9 ori, sau, în procente, cu 10%.

Scanned with OKEN Scanner

Calculele elementelor intermediare necesare pentru determinarea indicilor de grup se prezintă sistematizat în tabelul 21.4.

Tabelul 21.4

Produsul (j)	q_{oj}	p_{oj}	q_{1j}	p_{1j}	$q_{oj}p_{oj}$	$q_{1j}p_{oj}$	$q_{1j}p_{1j}$	$q_{oj}p_{1j}$
A	200	40	300	40	8000	12000	12000	8000
B	300	20	300	15	6000	6000	4500	4500
C	500	30	450	35	15000	13500	15750	17500
Total	-	-	-	-	29000	31500	32250	30000

Variaţia volumului fizic al tuturor produselor în perioada curentă faţă de perioada de bază se poate determina cu ajutorul indicelui agregat al volumului fizic de tip Ladpeyres dat de formula (21.11)pentru i=o,1 şi j=1,2,3, astfel:

$$I^{q}_{1/o} = \frac{\sum_{j=1}^{3} q_{1j}p_{oj}}{\sum_{j=1}^{3} q_{oj}p_{oj}} = \frac{31500}{29000} = 1,0862 \text{ sau } I^{q}_{1/o}\cdot 100-100=8,62\%$$

iar

$$\Delta^{q}_{1/o} = \sum_{j=1}^{3} q_{1j}p_{oj} - \sum_{j=1}^{3} q_{oj}p_{oj} = 31500-29000=2500 \text{ lei}$$

Rezultă că volumul fizic al celor trei produse(exprimate valoric) a crescut în perioada curentă faţă de perioada de bază cu 2500 lei sau de 1,0862 ori sau, în procente, cu 8,62%.

Această creştere a volumului fizic,exprimat valoric pe total produse se datorează,pe de o parte, faptului că volumul fizic al produsului A a crescut mai mult decît a scăzut cel al produsului C şi, într-o măsură mai mică şi în mod implicit, factorului de ponderare preţ, pe de altă parte.

Dinamica preţurilor pe fiecare produs şi pe total se determină cu ajutorul indicilor individuali (21.1) şi a celor agregat de tip Paasche (21.14):

-pentru produsul A:

$$i^{p_j}_{1/o} = \frac{p_1}{p_o} = \frac{4p}{4o} = 1 \quad \text{sau} \quad i^{p_j}_{1/o} \cdot 1oo - 1oo = o$$

iar

$$\Delta^{p_j}_{1/o} = p_1 - p_o = 4o - 4o = o$$

-pentru produsul B:

$$i^{p_j}_{1/o} = \frac{15}{2o} = o,75 \quad \text{sau} \quad i^{p_j}_{1/o} \cdot 1oo - 1oo = -25\%$$

iar

$$\Delta^{p_j}_{1/o} = 15 - 2o = -5 \text{ lei}$$

-pentru produsul C:

$$i^{p_j}_{1/o} = \frac{35}{3o} = 1,1667 \quad \text{sau} \quad i^{p_j}_{1/o} \cdot 1oo - 1oo = 16,67\%$$

iar

$$\Delta^{p_j}_{1/o} = 35 - 3o = 5 \text{ lei}$$

-pe total produse A,B şi C:

$$I^{p}_{1/o} = \frac{\sum_{j=1}^{3} p_{1j}q_{1j}}{\sum_{j=1}^{3} p_{oj}q_{1j}} = \frac{3225o}{315oo} = 1,o238 \quad \text{sau} \quad I^{p}_{1/o} \cdot 1oo - 1oo = 2,38\%$$

iar

$$\Delta^{p}_{1/o} = \sum_{j=1}^{3} p_{1j}q_{1j} - \sum_{j=1}^{3} p_{oj}q_{1j} = 3225o - 315oo = 75o \text{ lei}$$

Rezultă că variaria diferită a preţurilor celor trei produse în perioada curentă faţă de perioada de bază a determinat o creştere a preţurilor pe ansamblu de 1,o238 ori sau cu 2,38%. Această creştere a preţurilor a dus la o creştere a volumului valoric al producţiei cu 75o lei.

Ca şi în cazul indicelui volumului fizic(în general pentru toţi indicii sub formă agregat) valoarea indicelui agregat al preţurilor este influenţată şi de factorul de ponderare q_{1j}.

Variaţia volumului valoric al celor trei produse pe fiecare produs şi pe total se determină cu ajutorul indicilor individuali ai valorii (21.6) şi a celui agregat (21.15), astfel:

-pentru produsul A:

$$i^{v_j}_{1/o} = \frac{q_{1j}p_{1j}}{q_{oj}p_{oj}} = \frac{12000}{8000} = 1,5 \text{ sau } i^{v_j}_{1/o} \cdot 100 - 100 = 50\%$$

iar

$$\Delta^{v}_{1/o} = q_{1j}\,p_{1j} - q_{oj}p_{oj} = 12000 - 8000 = 4000 \text{ lei}$$

iar pe baza relaţiei 21.7 rezultă:

$$i^{v_j}_{1/o} = i^{q}_{1/o} \cdot i^{p}_{1/o} = 1,5 \cdot 1 = 1,5$$

Valoarea produsului A a crescut în perioada curentă faţă de perioada de bază cu 50% ca urmare a creşterii volumului fizic al acestui produs cu acelaşi procent şi a menţinerii constante a preţului.

-pentru produsul B

$$i^{v_j}_{1/o} = \frac{4500}{6000} = 0,75 \text{ sau } i^{v_j}_{1/o} \cdot 100 - 100 = -25\%$$

iar

$$\Delta^{v_j}_{1/o} = 4500 - 6000 = -1500 \text{ lei}$$

dar $$i^{v_j}_{1/o} = 1 \cdot 0,75 = 0,75$$

Aceasta înseamnă că valoarea produsului B a scăzut în perioada curentă faţă de perioada de bază cu 25% ca urmare a reducerii cu 25% a preţului acestui produs.

-pentru produsul C:

$$i^{v_j}_{1/0} = \frac{15750}{15000} = 1,05 \quad \text{sau} \quad i^{v_j}_{1/0} = 100-100 = 5\%$$

iar

$$\Delta^{v_j}_{1/0} = 15750-15000 = 750 \text{ lei}$$

dar

$$i^{v_j}_{1/0} = 0,9 \cdot 1,1667 = 1,05$$

Deci valoarea produsului C a crescut în perioada curentă față de perioada de bază cu 5% datorită creșterii cu 16,67% a prețurilor și reducerii cu 10% a volumului fizic al produsului luat în calcul.

-pe total produse

$$I^{v}_{1/0} = \frac{\sum_{j=1}^{3} q_{1j}p_{1j}}{\sum_{j=1}^{3} q_{0j}p_{0j}} = \frac{32250}{29000} = 1,112 \text{ sau } I^{v}_{1/0} \cdot 100-100=11,2\%$$

iar

$$\Delta^{v}_{1/0} = \sum_{j=1}^{3} q_{1j}p_{1j} - \sum_{j=1}^{3} q_{0j}p_{0j} = 32250-29000=3250 \text{ lei}$$

Rezultă că valoarea produselor A,B și C a crescut în perioada curentă față de perioada de bază cu 11,2% sau, în valoare absolută, cu 3250 lei.

Indicele valoric sub formă agregat mai poate fi determinat și cu ajutorul relației (21.18), astfel

$$I^{v}_{1/0} = I^{q}_{1/0} \cdot I^{p}_{1/0} = 1,0862 \cdot 1,0238 = 1,112$$

iar în mărimi absolute:

$$\Delta^{v}_{1/0} = \Delta^{q}_{1/0} + \Delta^{p}_{1/0} = 2500+750=3250 \text{ lei}$$

Indicele agregat al volumului fizic şi cel al preţurilor se pot calcula şi cu ajutorul formelor stabilite de Edgeworth (21.29) şi (21.3o) cele ale lui Drobisch (21.31) şi (21.32), pe baza tabelului 21.5.

Tabelul 21.5

Produsele(j)	q_{oj}	p_{oj}	q_{1j}	p_{1j}	p'_j	$q_{oj}p'_j$	$q_{1j}p'_j$	q'_j	q'_jp_{oj}	q'_jp_{1j}
A	2oo	4o	3oo	4o	8o	16ooo	24ooo	5oo	2ooo	2ooo
B	3oo	2o	3oo	15	35	1o5oo	1o5oo	6oo	12ooo	9ooo
C	5oo	3o	45o	35	65	325oo	2925o	95o	285oo	3325o
Total	-	-	-	-	-	59ooo	6375o	-	6o5oo	6225o

Indicii de tip Edgeworth:

$$I^{q}_{1/o} = \frac{\sum_{j=1}^{3} q_{1j}(p_{oj}+p_{1j})}{\sum_{j=1}^{3} q_{oj}(p_{oj}+p_{1j})} = \frac{\sum_{j=1}^{3} q_{1j}p'_j}{\sum_{j=1}^{3} q_{oj}p'_j} = \frac{6375o}{59ooo} = 1,o8o5$$

şi

$$I^{p}_{1/o} = \frac{\sum_{j=1}^{3} p_{1j}(q_{oj}+q_{1j})}{\sum_{j=1}^{3} p_{oj}(q_{oj}+q_{1j})} = \frac{\sum_{j=1}^{3} p_{1j}q'_j}{\sum_{j=1}^{3} p_{oj}q'_j} = \frac{6225o}{6o5oo} = 1,o289$$

Pentru a calcula indicile volumului fizic şi al preţurilor de tip Drobisch va trebui să determinăm mai întîi indicele agregat de tip Paasche pentru volumul fizic şi indicele agregat de tip Laspeyres pentru preţuri, astfel:

$$I^{q}_{1/o} = \frac{\sum_{j=1}^{3} q_{1j}p_{1j}}{\sum_{j=1}^{3} q_{oj}p_{1j}} = \frac{3225o}{3ooo5} = 1,o75$$

şi $$I^p_{1/o} = \frac{\sum_{j=1}^{3} q_{oj}p_{ij}}{\sum_{j=1}^{3} q_{oj}p_{oj}} = \frac{30000}{29000} = 1,0344$$

Observăm că indicele volumului fizic şi cel al preţurilor iau valori diferite atunci cînd se utilizează sisteme de ponderare diferite.

Indicii de tip Drobisch:

$$I^q_{1/o} = \left(\frac{\sum_{j=1}^{3} q_{ij}p_{oj}}{\sum_{j=1}^{3} q_{oj}p_{oj}} + \frac{\sum_{j=1}^{3} q_{1j}p_{1j}}{\sum_{j=1}^{3} q_{oj}p_{1j}} \right) : 2 =$$

$$= (1,0862 + 1,075) : 2 = 1,0806$$

şi

$$I^p_{1/o} = \left(\frac{\sum_{j=1}^{3} q_{oj}p_{tj}}{\sum_{j=1}^{3} q_{oj}p_{oj}} + \frac{\sum_{j=1}^{3} q_{1j}p_{1j}}{\sum_{j=1}^{3} q_{1j}p_{oj}} \right) : 2 =$$

$$= (1,0344 + 1,0238) : 2 = 1,0291$$

Se observă că indicii de tip Edgeworth şi cei de tip Paasche au valori apropiate pentru aceleaşi fenomene, iar aceste valori se situează între valorile indicilor de tip Laspeyres şi celor de tip Paasche.

Indici sub formă de agregat se pot construi şi pentru mai mult de doi factori.

In cazul a trei factori x, y şi z indicele agregat

care oglindeşte influenţa tuturor celor trei facteri se va calcula după formulele:

-cu bază fixă:

$$I^{xyz}_{i/o} = \frac{\sum_j x_{ij}\, y_{ij}\, z_{ij}}{\sum_j x_{oj}\, y_{oj}\, z_{oj}} \qquad \begin{matrix} i = o,1,2,\ldots, k \\ j = o, 1,2,\ldots,n \end{matrix} \qquad (21.33)$$

şi

$$\Delta^{xyz}_{i/o} = \sum_j x_{ij}\, y_{ij}\, z_{ij} - \sum x_{oj}\, y_{oj}\, z_{oj}$$

-cu bază mobilă:

$$I^{xyz}_{i/i-1} = \frac{\sum_j x_{ij}\, y_{ij}\, z_{ij}}{\sum_j x_{i-1,j} y_{i-1,j} z_{i-1,j}} \qquad (21.34)$$

iar

$$\Delta^{xyz}_{i/i-1} = \sum_j x_{ij} y_{ij} z_{ij} - \sum_j x_{i-1,j} y_{i-1,j} z_{i-1,j}$$

In cazul luării în calcul a patru facteri x,y,z,v indicii sub formă agregat se vor calcula după formulele.

-cu bază fixă:

$$I^{xyzv}_{i/o} = \frac{\sum_j x_{ij} y_{ij} z_{ij} v_{ij}}{\sum_j x_{oj} y_{oj} z_{oj} v_{oj}} \qquad (21.35)$$

iar

$$\Delta^{xyzv}_{i/o} = \sum_j x_{ij} y_{ij} z_{ij} v_{ij} - \sum_j x_{oj} y_{oj} z_{oj} v_{oj}$$

-cu bază mobilă:

$$I^{xyzv}_{i/i-1} = \frac{\sum_j x_{ij} y_{ij} z_{ij} v_{ij}}{\sum_j x_{i-1,j} y_{i-1,j} z_{i-1,j} v_{i-1,j}} \qquad (21.36)$$

iar

$$\Delta^{xyzv}_{i/i-1} = \sum_j x_{ij} y_{ij} z_{ij} v_{ij} - \sum_j x_{i-1,j} y_{i-1,j} z_{i-1,j} v_{i-1,j}$$

21.3.2.2. Indicele mediu aritmetic ponderat

O formă transformată a indecelui agregat o constituie indicele mediu aritmetic ponderat, care se calculează ca o medie aritmetică a indicilor individuali ai fenomenului studiat ponderaţi cu elementele de la numitorul formulei indicelui agregat corespunzător.

Astfel, atunci cînd se cunosc indicii individuali ai volumului fizic pentru un grup de produse şi valorile acestor produse din perioada de bază, dinamica volumului fizic al producţiei grupei de produse se poate calcula după formula:

$$I^{q}_{i/o} = \frac{\sum_j i^{q}_{i/o} \cdot q_{oj} p_{oj}}{\sum_j q_{oj} p_{oj}} \qquad (21.37)$$

Dacă se substituie în formula (21.37) $i^{q}_{i/o} = \frac{q_{ij}}{q_{oj}}$ rezultă:

$$I^{q}_{i/o} = \frac{\sum_j \frac{q_{ij}}{q_{oj}} \cdot q_{oj} p_{oj}}{\sum_j q_{oj} p_{oj}} = \frac{\sum_j q_{ij} p_{oj}}{\sum_j q_{oj} p_{oj}} \qquad (21.38)$$

Relaţia (21.38) arată că indicele mediu aritmetic ponderat este egal cu indicele agregat corespunzător.

Pentru exemplificarea calculului şi analizei acestui indice să considerăm că se cunosc indicii individuali ai

volumului fizic pentru cele trei produse din exemplul anterior (tabelul 21.3) şi ce mai cunosc valorile celor trei produse din perioada de bază exprimate în preţurile perioadei (tabelul 21.6).

Tabelul 21.6

Produsele	$i_{i/o}^{q_j}$	$q_{oj}p_{oj}$	$i_{i/o}^{q_j} q_{oj}p_{oj}$
A	1,5	8000	12000
B	1,0	6000	6000
C	0,9	15000	13500
Total	-	29000	31500

Substituind datele din tabelul 21.6 în formula (21.37) se obţine:

$$I_{i/o}^{q} = \frac{31500}{29000} = 1,0862$$

Se observă că s-a obţinut aceeaşi valoare a indicelui de grup al volumului fizic ca şi în cazul utilizării formulei indicelui agregat de tip Laspeyres.

21.3.2.3. Indicele mediu armonic ponderat

Acest indice se utilizează în activitatea statistică atunci cînd se cunosc indicii individuali ai fenomenului cercetat şi elementele de la numărătorul formulei indicelui agregat corespunzător. Indicele de grup mediu armonic ponderat se calculează ca o medie armonică ponderată a indicilor individuali ai fenomenului cercetat ponderaţi cu elementele de la numărătorul formulei indicelui agregat corespunzător. Se aplică în practica statistică de regulă pentru determinarea dinamicii preţurilor. Indicele mediu armonic ponderat al preţurilor se calculează pe baza formulei:

$$I_{i/o}^{p} = \frac{\sum_j p_{ij} q_{ij}}{\sum_j \frac{1}{i_{i/o}^{p}} p_{ij} q_{ij}} \qquad (21.39)$$

Dacă în formula (21.39) se substituie $i^{p}_{1/o} = \frac{p_{1j}}{p_{oj}}$ rezultă:

$$I^{p}_{1/o} = \frac{\sum_j p_{1j}q_{1j}}{\sum_j \frac{1}{\frac{p_{1j}}{p_{oj}}} p_{1j}q_{1j}} = \frac{\sum_j p_{1j}q_{1j}}{\sum_j p_{oj}q_{1j}} \qquad (21.4o)$$

Din relaţia (21.4o) rezultă că indicele mediu armonic ponderat este echivalent cu indicele agregat coresponzător.

Să considerăm, pentru exemplificare, că pentru cele trei produse din exemplul anterior se cunosc indicii individuali ai preţurilor şi valoarea produselor din perioada curentă exprimată în preţurile acestei perioade(tabelul 21.7).

Tabelul 21.7.

Produsul	$i^{p_j}_{1/o}$	$q_{1j}p_{1j}$	$\frac{q_{1j}p_{1j}}{i^{p_j}_{1/o}}$
A	1	12ooo	12ooo
B	o,75	45oo	6ooo
C	1,1667	15750	135oo
Total	-	3225o	315oo

Înlocuind în formula (21.39) datele din tabelul 21.7 obţinem:

$$I^{p}_{1/o} = \frac{3225o}{315oo} = 1,o238$$

Se verifică şi în acest caz relaţia de egalitate dintre indicele agregat al preţurilor de tip Paasche şi indicele mediu armonic ponderat al preţurilor.

21.3.2.4.Indicele mediu geometric

W.S.Jevons într-un studiu binecunoscut "A serious

fall in the value of gold ascertained", publicat în 1863 calculează pentru prima dată indici cu ajutorul formulei mediei geometrice simple din indici individuali, prin extragerea rădăcinii de ordinul "n" din produsul a"n" indici simpli, după formula:

$$I^{p}_{i/o} = \sqrt[n]{\frac{p_{i1}}{p_{o1}} \cdot \frac{p_{i2}}{p_{o2}} \cdot \ldots \cdot \frac{p_{ij}}{p_{oj}} \cdot \ldots \cdot \frac{p_{in}}{p_{on}}} \qquad (21.41)$$

Calculul practic al unui astfel de indice se face prin logaritmarea expresiei (21.41), obținîndu-se:

$$\log I^{p}_{i/o} = \frac{\log \frac{p_{i1}}{p_{o1}} + \log \frac{p_{i2}}{p_{o2}} + \ldots + \log \frac{p_{ij}}{p_{oj}} + \ldots \log \frac{p_{in}}{p_{on}}}{n}$$

Jevons a studiat problema scăderii prețului aurului ca urmare a descoperirii de noi zăcăminte în California și Australia în jurul anului 1850. El a luat în calcul prețurile a 39 de produse din anii 1845-1862, considerînd ca prețuri de bază pentru fiecare produs, media aritmetică a prețurilor din perioada 1845-1850. Apoi a împărțit prețurile, corespunzătoare pentru fiecare din cei 18 ani ai perioadei la nivelul lor mediu de bază, obținînd astfel indicii individuali ai prețurilor, pe baza cărora a determinat nivelul lor mediu cu ajutorul mediei geometrice simple.

Deficiența principală a indicelui lui Jevons constă în faptul că nu a luat în considerare volumul fizic al produselor drept pondere.

Cu toate acestea metoda lui Jevons a stat mai tîrziu la baza faimoasei formule a indicelui ideal al prețurilor elaborată de către C.M.Walsh în lucrarea sa "Measurements of General Exchange Value" într-o notă de subsol al unei pagini în anul 1901. Aceeași formulă este publicată de A.C.Pigou în anul 1912 în lucrarea "Wealth and Welfare" la pagina 46, dar nu extrage rădăcina pătrată din produs. Această deficiență este înlăturată de autor în anul 1920 în lucrarea "Economics of Welfare" la pagina 78.

Problema construirii unui indice ideal este preluată şi analizată sub multiple aspecte de către statisticianul american Irving Fisher care o prezintă într-o comunicare făcută la American Statistical Association în anul 1921.

Indicele mediu geometric îmbină calităţile indicelui agregat de tip Laspeyres cu cele ale indicelui de tip Paasche şi se calculează ca o medie geometrică simplă a celor două tipuri fundamentale ale indicelui agregat bifactorial, după formulele:

-în cazul volumului fizic:

$$I^{q}_{i/0}=\sqrt{\frac{\sum_{j} q_{ij}p_{oj}}{\sum_{j} q_{oj}p_{oj}}\cdot\frac{\sum_{j} q_{ij}p_{ij}}{\sum_{j} q_{oj}p_{ij}}} \qquad (21.42)$$

-în cazul preţurilor:

$$I^{p}_{i/0}=\sqrt{\frac{\sum_{j} p_{ij}q_{oj}}{\sum_{j} p_{oj}q_{oj}}\cdot\frac{\sum_{j} p_{ij}q_{ij}}{\sum_{j} p_{oj}q_{ij}}} \qquad (21.43)$$

Pentru exemplificarea calculului şi analizei indicelui mediu geometric să considerăm datele din tabelul 21.4.

Indicele mediu geometric al preţurilor celor trei produse se obţine pe baza formulei (21.43), astfel:

$$I^{p}_{i/0}=\sqrt{\frac{\sum_{j=1}^{3} p_{ij}q_{oj}}{\sum_{j=1}^{3} p_{oj}q_{oj}}\cdot\frac{\sum_{j=1}^{3} p_{ij}q_{ij}}{\sum_{j=1}^{3} p_{oj}q_{ij}}}=\sqrt{\frac{30000}{29000}\cdot\frac{32250}{31500}}=$$

$$=\sqrt{1,0344.1,0238}\simeq 1,029$$

Valoarea indicelui mediu geometric ponderat al preţurilor în mod logic este cuprinsă între valoarea indicelui agregat al preţurilor de tip Laspeyres şi valoarea indicelui agregat al preţurilor de tip Paasche al căror medie este.

21.3.2.5. Indici calculaţi ca raport între două medii aritmetice ponderate

Sistemul indicilor calculaţi ca raport între două medii aritmetice ponderate este compus din trei categorii de indici:

- indici bifactoriali;
- indici unifactoriali cu structură fixă;
- indici unifactoriali cu variaţii ale structurii.

Indicii bifactoriali măsoară variaţia nivelului mediu al unui fenomen complex în două perioade de timp diferite sau al aceluiaşi fenomen, în acelaşi timp, dar situat în spaţii diferite. Deoarece în calculul nivelului mediu intră doi factori - nivelurile individuale ale fenomenului studiat (x_j) şi structura colectivităţii $\left(\frac{f_{ij}}{\sum_j f_{ij}}\right)$ - indicele bifactorial va arăta influenţa concomitentă a două categorii de factori şi se va calcula după formula:

$$I^{xf}_{i/o} = \frac{\bar{x}_i}{\bar{x}_o} = \frac{\sum_j x_{ij} f_{ij}}{\sum_j f_{ij}} : \frac{\sum_j x_{oj} f_{oj}}{\sum_j f_{oj}} \qquad (21.35)$$

Indicele bifactorial se poate descompune în doi indici unifactoriali unul care să arate numai variaţia caracteristicii (indicele cu structură fixă) şi altul care să arate numai influenţa variaţiei structurii colectivităţii (indicele variaţiei structurii).

Substituind în formula (21.35) elementul de structură din perioada de bază f_{oj} cu cel din perioada curentă f_{ij} se obţine formula uzuală a indicelui cu structură fixă:

$$I^{x}_{i/o} = \frac{\sum_j x_{ij} f_{ij}}{\sum_j f_{ij}} : \frac{\sum_j x_{oj} f_{ij}}{\sum_j f_{ij}} = \frac{\sum_j x_{ij} f_{ij}}{\sum_j x_{oj} f_{ij}} \qquad (21.36)$$

Din relaţia (21.36) se observă că indicele cu structură fixă are forma unui indice agregat de tip Paasche cu toate calităţile şi deficienţele acestuia.

Dacă în formula (21.35) se menţine constant nivelul caracteristicii atunci se va obţine un indice unifactorial care arată numai influenţa variaţiei caracteristicii şi poartă denumirea de indice al variaţiei structurii:

$$I^{f}_{i/o} = \frac{\sum_j x_{oj} f_{ij}}{\sum_j f_{ij}} : \frac{\sum_j x_{oj} f_{oj}}{\sum_j f_{oj}} \qquad (21.37)$$

Între indicele bifactorial (21.35) şi indicii unifactoriali (21.36) şi (21.37) există o relaţie de sistem, bine definită şi anume: produsul dintre cei doi indici unifactoriali este egal cu indicele bifactorial.

Deci

$$I^{xf}_{i/o} = I^{x}_{i/o} \cdot I^{f}_{i/o} \qquad (21.38)$$

Substituind în (21.38) cei doi indici unifactoriali cu expresiile lor, rezultă:

$$\frac{\sum_j x_{ij} f_{ij}}{\sum_j x_{oj} f_{ij}} \cdot \frac{\sum_j x_{oj} f_{ij}}{\sum_j f_{ij}} : \frac{\sum_j x_{oj} f_{oj}}{\sum_j f_{oj}} =$$

$$= \frac{\sum_j x_{ij} f_{ij}}{\sum_j f_{ij}} : \frac{\sum_j x_{oj} f_{oj}}{\sum_j f_{oj}}$$

Din relaţia (21.38) rezultă alte două relaţii utile în activitatea statistică:

a.Indicele cu structură fixă se poate calcula şi ca un raport între indicele bifactorial şi indicele variaţiei structurii:

$$I^{x}_{i/o} = I^{xf}_{i/o} : I^{f}_{i/o}$$

b.Indicele variaţiei structurii este egal cu raportul dintre indicele bifactorial şi indicele cu structură fixă:

$$I^{f}_{i/o} = I^{xf}_{i/o} : I^{x}_{i/o}$$

Dacă în sistemul indicilor calculaţi ca raport între două medii aritmetice ponderate se substituie factorul $\frac{f_{ij}}{\sum f_{ij}} = G_{ij}$ atunci se obţine forma agregat a celor trei indici, astfel:

$$I^{xf}_{i/o} = \frac{\sum_j x_{ij}f_{ij}}{\sum_j f_{ij}} : \frac{\sum_j x_{oj}f_{oj}}{\sum_j f_{oj}} = \frac{\sum_j x_{ij}G_{ij}}{\sum_j x_{oj}G_{oj}} \qquad (21.39)$$

şi

$$I^{x}_{i/o} = \frac{\sum_j x_{ij}f_{ij}}{\sum_j f_{ij}} : \frac{\sum_j x_{oj}f_{ij}}{\sum_j f_{ij}} = \frac{\sum_j x_{ij}G_{ij}}{\sum_j x_{oj}G_{ij}} \qquad (21.40)$$

$$I^{f}_{i/o} = \frac{\sum_j x_{oj}f_{ij}}{\sum_j f_{ij}} : \frac{\sum_j x_{oj}f_{oj}}{\sum_j f_{oj}} = \frac{\sum_j x_{oj}G_{ij}}{\sum_j x_{oj}G_{oj}} \qquad (21.41)$$

Relaţia (21.38) se verifică şi în acest caz:

$$\frac{\sum_j x_{ij}G_{ij}}{\sum_j x_{oj}G_{oj}} = \frac{\sum_j x_{ij}G_{ij}}{\sum_j x_{oj}G_{ij}} \cdot \frac{\sum_j x_{oj}G_{ij}}{\sum_j x_{oj}G_{oj}} \qquad (21.42)$$

Să considerăm, pentru exemplificare, situaţia producţiei dintr-o centrală industrială formată din trei întreprinderi. (tabelul 21.8)

Tabelul 21.8

Întreprinderea	Perioada de bază		Perioada curentă		W_o	W_1	$i^W_{1/o}$	W_oT_1
	Volumul valoric al producţiei (mii lei)	Timpul de muncă cheltuit (mii om-zile)	Volumul valoric al producţiei (mii lei)	Timpul de muncă cheltuit (mii om-zile)				
o	1	2	3	4	5	6	7	8
A	5000	20	8000	25	250	320	1,28	6250
B	10000	40	9984	32	250	312	1,248	8000
C	4950	22	3000	10	225	300	1,333	2250
Total	19950	82	20984	67	243,3	313,2	1,287	16500

Variaţia productivităţii zilnice a muncii pe fiecare întreprindere în perioada curentă faţă de perioada de bază, este dată de indicii individuali ai productivităţii muncii (col.7 a tabelului 21.8). Se observă că a avut loc o creştere a productivităţii muncii pe fiecare întreprindere din cadrul centralei şi anume o creştere cu 33,3% în întreprinderea C, cu 28% în întreprinderea A şi cu 24,8% în întreprinderea B.

La nivel de centrală dinamica productivităţii zilnice a muncii este dată de indicii de grup ai productivităţii muncii.

Astfel, variaţia productivităţii zilnice a muncii la nivel de centrală datorată influenţei tuturor factorilor care au acţionat asupra ei se determină cu ajutorul indicelui bifactorial al productivităţii muncii (21.35).

Substituind datele din tabelul 21.8 în formula 21.35 rezultă:

$$I^{W,T}_{1/o} = \frac{\sum W_1T_1}{\sum T_1} : \frac{\sum W_oT_o}{\sum T_o} = \frac{20984}{67} : \frac{19950}{82} = 1,287$$

$I^{W,T}_{1/o} = 1,287$ înseamnă că productivitatea zilnică a muncii pe centrală a crescut cu 28,7% în perioada curentă faţă de perioada de bază ca urmare a creşterii productivităţii

muncii pe fiecare întreprindere şi a variaţiei structurii centralei pe întreprinderi din punct de vedere al timpului de muncă cheltuit pentru realizarea producţiei.

Influenţa separată a factorilor interni ai producţiei (mai buna organizare a procesului de producţie în întreprinderi, creşterea calificării muncitorilor, introducerea tehnicii noi etc.) este dată de indicele unifactorial cu structură fixă al productivităţii muncii (21.36) care, aplicat la exemplul nostru, va fi:

$$I^{W}_{1/o} = \frac{\sum W_1 T_1}{\sum W_o T_1} = \frac{20.984}{16.500} = 1,272$$

Deci cea mai mare parte a creşterii productivităţii muncii la nivel de centrală, şi anume 27,2%, s-a obţinut pe seama factorilor interni ai producţiei.

Cea de a doua categorie de factori care au acţionat asupra dinamicii productivităţii muncii pe centrală (factorii structurali sau "combinarea socială a procesului de producţie") este dată de indicele unifactorial cu variaţie a structurii (21.37). Substituind datele din tabelul 21.8 în formula (21.37) rezultă:

$$I^{T}_{1/o} = \frac{\sum W_o T_1}{\sum T_1} : \frac{\sum W_o T_o}{\sum T_o} = \frac{16500}{67} : \frac{19950}{82} = 1,012$$

Aceasta înseamnă că factorii structurali au avut o influenţă mai mică asupra dinamicii productivităţii muncii pe centrală şi anume ei au determinat o creştere a acestuia de 1,2%.

Relaţia de sistem (21.38) se verifică astfel:

$$1,287 \simeq 1,272 \cdot 1,012$$

Desigur acelaşi rezultat s-ar fi obţinut şi dacă s-ar fi utilizat indicii sub formă agregat pentru calculul dinamicii productivităţii muncii la nivel de centrală.

Calculele intermediare se sistematizează cu ajutorul tabelului 21.9.

Tabelul 21.9

Intreprinderea	T	T_1	W_o	W_1	$G_o=\frac{T_o}{\sum T_1}$	$G_1=\frac{T_1}{\sum T_1}$	W_oG_o	W_oG_1	W_1G_1
A	2o	25	25o	32o	o,2439	o,3731	6o,975	93,275	119,392
B	4o	32	25o	312	o,4878	o,4776	121,95o	119,4oo	149,o11
C	22	1o	225	3oo	o,2683	o,1493	6o,367	33,593	44,79o
Total	82	67	243,3	313,2	1,oooo	1,oooo	243,292	246,268	313,193

Particularizînd relaţia (21.39) pentru dinamica productivităţii muncii rezultă formula indicelui agregat bifactorial al productivităţii muncii:

$$I^{W,G}_{1/o}=\frac{\sum W_1G_1}{\sum W_oG_o}=\frac{313,193}{243,292}=1,287$$

Indicele agregat cu structură fixă al productivităţii muncii va fi:

$$I^{W}_{1/o}=\frac{\sum W_1G_1}{\sum W_oG_1}=\frac{313,193}{246,268}=1,272$$

Indicele agregat al productivităţii muncii cu variaţia a structurii este dat de formula:

$$I^{G}_{1/o}=\frac{\sum W_oG_1}{\sum W_oG_o}=\frac{246,268}{243,292}=1,o12$$

Observăm că s-au obţinut aceleaşi rezultate ca şi în cazul utilizării indicilor productivităţii muncii calculaţi din mărimi medii.

Utilizarea formei agregat a indicilor productivităţii muncii (şi în general a oricărui indice calculat din mărimi medii) dă posibilitatea determinării contribuţiei fiecărei întreprinderi la formarea nivelului mediu al productivităţii muncii, pe centrală. Astfel în perioada curentă productivitatea medie zilnică a muncii pe centrală a fost de 313,19 lei pe un muncitor; la care întreprinderea A a contribuit cu 119,392 lei, întreprinderea B cu 149,oll lei şi în-

treprinderea C cu 44,79 lei.

Creşterea productivităţii medii zilnice a muncii la nivel de centrală, în perioada curentă faţă de perioada de bază este dată de diferenţa dintre numărătorul şi numitorul indicelui agregat bifactorial, astfel:

$$\Delta \frac{W,G}{I/_o} = \sum W_1 G_1 - \sum W_o G_o = 313,193-243,292= 69,90 \text{ lei/om-zi}$$

Aceasta înseamnă că productivitatea medie zilnică pe un muncitor la nivel de centrală a crescut în perioada curentă faţă de perioada de bază cu 69,90 lei ca urmare a influenţei ambelor categorii de factori.

Influenţa separată a celor două categorii de factori va fi:

-influenţa creşterii productivităţii muncii pe fiecare întreprindere a centralei în perioada curentă faţă de perioada de bază, dată de diferenţa dintre numărătorul şi numitorul formulei indicelui cu structură fixă, este

$$\Delta \frac{W}{I/_o} = \sum W_1 G_1 - \sum W_o G_1 = 313,193-246,268 = 66,93 \text{ lei/om-zi}$$

ceea ce arată că factorii interni ai procesului de producţii din întreprinderi au avut o influenţă favorabilă asupra creşterii productivităţii zilnice a muncii la nivel de centrală.

-influenţa factorilor structurali va fi dată de diferenţa dintre numărătorul şi numitorul formulei indicelui cu variaţii ale structurii, astfel:

$$\Delta \frac{G}{I/_o} = \sum W_o G_1 - \sum W_o G_o = 246,268-243,292=2,98 \text{ lei}$$

Desigur, că între influenţele factorilor productivităţii muncii, exprimate în mărimi absolute, se va verifica relaţia de sistem:

$$\Delta \frac{W,G}{I/_o} = \Delta \frac{W}{I/_o} + \Delta \frac{G}{I/_o}$$

iar în exemplul nostru vom avea:

$$69,9 \simeq 66,92 + 2,97$$

Indicii calculaţi prin raportarea a două medii aritmetice ponderate se pot calcula pentru studierea dinamicii oricăror fenomene. Mai des utilizaţi în practică sînt indicii productivităţii muncii, ai costurilor de producţie, ai consumurilor specifice, ai nivelului relativ al cheltuielilor de circulaţie (în comerţ), ai recoltei medii la hectar, ai mortalităţii populaţiei etc.

21.3. Erorile de măsurare în construcţia indicilor

In funcţie de cauzele care determină apariţia erorilor de măsurare în construcţia indicilor, se disting trei componente ale lor:

- eroare dată de formulă;
- eroare de sondaj;
- eroare de omogenitate.

Eroarea dată de formulă se produce la calculul oricărui indice agregat al volumului fizic sau al preţurilor deoarece nu există o formulă general valabilă. Formulele lui Laspeyres şi Paasche sînt la fel de bune şi deci la fel de rele. Aceasta înseamnă că nu există argumente ştiinţifice concludente în favoarea alegerii uneia sau a celeilalte.

Datorită faptului că în calculele concrete ale indicilor nu este posibil sau nu este eficient să se ia în calcul toate produsele ci numai o parte a lor, aleasă selectiv, rezultatele obţinute cu ajutorul indicilor vor fi afectate şi de erori de sondaj care se pot măsura şi interpreta probabilistic.

Pentru a efectua o comparaţie binară exactă este necesar să se ia în calcul elementele perioadei de bază şi ale perioadei curente. Se ştie însă că în practică orice produs are o anumită durată a existenţei lui. Aceasta face ca anumite produse care existau în perioada de bază să nu mai existe în perioada de comparat sau în intervalul dintre cele două perioade să fi apărut noi produse. Deci oricum am proceda comparaţia binară va fi afectată de o anumită eroare. Astfel, dacă am lua în calcul toate produsele care există în fiecare din cele două perioade, atunci acestea n-ar fi în totalitate comparabile. Dacă vom lua în calcul numai produsele care au existat în ambele perioade atunci înseamnă că vom exclude anumite produse "unice" pentru o anumită perioadă. Această eroare poartă denumirea de <u>eroare de omogenitate</u> şi nu se cunosc pînă în prezent metode potrivite pentru măsurarea ei.

BIBLIOGRAFIE GENERALA SELECTIVA

1. Alessandro, B. - Compendiu general de statistică, Graza Ciesler, 1871, 296 p.

2. Allen, R.G.D. - Macro-Economic Theory, St. Martin's Press Inc., New York, 1967.

3. Alterman, Hyman - Numbers at work: the story and science of statistics, F.W. Taylor, New York, Harcour-Brace & World, 1966, 280 p.

4. Bărbat, Alexandru - "Teoria statisticii sociale", Ed. did. şi ped., Bucureşti, 1972.

5. Beals, Ralph E. - "Statistics for economists: an introduction, Chicago, Rand McNally, 1972, 420 p.

6. Berechet, Ştefan - Curs de finanţe şi statistică, Univ. Mihăileană Iaşi, Fac.de Drept, 1936, 461 p.

7. Bertaud , Marcel - "Introduction à la statistiques. Cours programmé", Les Presses de l'Université de Montréal, Montréal, 1973 .

8. Biji, M., Haşigan,D., Pintilie,L., Stoichiţă, I., Tövissi,L. - "Curs de statistică teoretică", Ed. did. şi ped., Bucureşti, 1963.

9. Blalock, Hubert M. - "Social Statistics", 2nd edition, McGraw-Hill Book Company, New York, 1972.

10. Bowley, Arthur - "Elements de statistique", Paris, 1929.

11. Brennan, Michael Joseph - "Theory of Economic Statistics", 2d ed, Englewood Cliffs, N.J., Prentice Hall, 1970, 437 p.

12. Bryant, Edward Clark - Statistical analysis, New York, McGraw Hill, 1966, 321 p.

13. Calot,Gerard - Cours de statistique descriptive, Dunod, Paris, 1969.

14. Carlson, Roger A. - Statistics, San Francisco, Holden-Day, 1973, 393 p.

15. Chakravati, I.M., Laha,R.G. and Roy,Jogabrata - Handbook of Methods of Applied Statistics, Calcuta Research and Training School, Indian Statistical Institute, 1957, Vol. I: General Metods.
Vol. II: Applications.

16. Chao,Lincoln L. - Statistics:methods and analyses, New York, McGraw-Hill, 1969, 512 p.

17. Chase, Clinton I. - Elementary statistical procedures,New York, McGraw-Hill, 1967, 245 p.

18. Chou, Ya-Lun - Statistical analysis with business and economic applications, New York, Holt, Rinehart and Winston, 1969, 794 p.

19. Chou, Ya-Lun - Probabily and statistics for decision making, New York, Holt,Rinehart and Winston,1972, 623 p.

20. Chris, F. (Ed.) - Measurement in Economics, Palo Alto, California, Stanford University Press, 1963.

21. Cihoschi, S. - Curs de statistică generală, Academia de Inalte Studii Comerciale şi Industriale, Bucureşti, 1922, lito.

22. Clark, Charles Tallifero - Statistical Methods for business decisions, Cincinnati, Southwestern Pub.Co.,1969, XII, 750 p.

23. Clarke, G.M. and D. Cooke - A Basic Course in Statistics, Edward Arnold Publishers Ltd., London, 1978.

24. Clelland, Richard C. - Basic Statistics with business applications, New York, Wiley, 1966, 708 p.

25. Conover, W.J. - Practical Nonparametric Statistics, John Wiley and Sons Inc., New York,1971, 462 p.

26. Copper, B.E. - Statistics for Experimentalists, Pergamon Press, Oxford, 1975.

27. Cristureanu, T., Ivănescu, I., Negura, I., Stoianovici, G. - Sistemul informational si calculul statistico-economic în comertul exterior, Ed. ştiintifică, Bucureşti, 1974.

28. Croxton, Frederick E. - Elementary Statistics, Dover Publications, Inc., New York, 1959.

29. Crow, Edwin, Davis Frances, Maxfield Margaret - Statistics Manual, Dover Publications Inc., New York, 1960.

30. Davis, G.R. and D. Yoder - Business Statistics, 2nd ed., John Wiley & Sons, Inc., New York, 1941.

31. David, Herbert Aron - Order Statistics, New York, J. Wiley, 1970, 272 p.

32. Darmois, Georges - Statistique et applications, Paris, 1957.

33. Debaty, Pol - La statistique parametrique, Editions Universitaires, Paris, 1967.

34. Decusară, E.C. - Elemente de statistică teoretică si aplicată, Ed. Inst. Central de statistică, Bucureşti, 1943.

35. Dodulescu, P. - Notiuni de statistică generală si demografică, Oradea-Mare, 1924, XIV, 240 p.

36. Dixon, Wilfrid Joseph - Introduction to statistical analysis, New York, McGraw-Hill, 1969, 638 p.

37. Dolguşevski, P.G., Kozlov V.S., Polusin P.I., Erlih M. - Obşcaia teoria statistiki, izd. Statistika, Moskva, 1967.

38. Dollar, Charles M. - Historian's guide to statistics ... quantitative analysis and historical research, New York, Holt, Rinehart and Winston, 1971, 332 p.

39. Donda, Arno, Herrde Eva, Kuhn Olfried, Struck Richard - Statistik, Verlag Die Wirtschaft, Berlin, 1972.

40. Drèze, J. - Cours d'analyse statistique, l'Université Catolic de Louvain, Louvain, 1977.

41. Dubois, Jean - Statistiatique descriptive en science économique, Paris, Dunod, 1970, XIV, 408 p.

42. Duncan, Otis Dudley - "Statistical geography : problems in analysing areal data, Free Press, 1961, 191 p.

43. Edwards, Allen L. - Statistical Methods, 2nd edition, Holt, Rinehart and Winston, Inc., New York, 1967.

44. Ferguson, C.E., Gould, J.P. - Microeconomic Theory, Richard D. Irwin, Inc., Homewood, 4th edition, 1975.

45. Fisher, Sir Ronald Aylmer - Statistical methods for research workar, New York, Hafner, 1954.

46. Frank, Charles Raphael - Statistics and econometrics, New York, Holt, Rinehart and Wiston, 1971, 400 p.

47. Freund, John E. - Modern Elementary Statistics, 3rd edition, Prentice-Hall Inc., Englewood Cliffs, 1967.

48. Georgescu-Roegen, N. Şt. - Metoda statistică. Elemente de statistică matematică, Imprimeria Națională, București, 1933.

49. Georgescu-Roegen, Nicholas - Analytical Economics, Harvard University Press, Cambridge, 1967.

50. Georgescu-Roegen, Nicholas - "The Entropy Law and the Economic Process", Harvard University Press, Cambridge, 1971.

51. Gini, Corrado - La logica nella statistica, Editore Boringhieri, Torino, 1962.

52. Girault, M. - Eléments de Méthodologie statistique, Dunod, Paris, 1967.

53. Grais, Bernard - Technique statistique, vol. I Statostique descriptive; vol. II Méthodes statistiques, Dunod editeur, Paris, 1977.

54. Gromiko, G.L. - Statistika, izd. Moskovskogo Universiteta, Moskva, 1976.

69. Koopmans, T.C. - Statistical Inference in Dinamic Economic Models, Wiley, New York, 1950.

70. Korin, Basil P. - Statistical Concepts for the Social Sciences, Winthrop Publishers, Inc., Cambridge, 1975.

71. Kovačka Milan, Kontšekova Olga - Štatistiké Metódy, tretie nezmenené vydanie, Nakladatelstvo ALFA, Bratislava, 1969.

72. Kraft, Charles Hall - A nonparametric introduction to statistics, New York, Macmillan, 1968, 342 p.

73. Kusnets, Simion - Modern Economic Growth, Yale University Press, New Haven and London, sixth printing, 1973.

74. Labrousse, Christian - Statistique. Exercices Corrigés, 2 vol., Dunod, Paris, 1965.

75. Lange, J. - Éléments de technique statistique, Dunod, Paris, 1968.

76. Langley, Russell- Practical Statistics, Dover Publication, Inc., New York, 1970.

77. Leabo, Dick Albert - Basic statistics, Homewood, Ill, R.D. Irwin. 1972.

78. Lebart, Ludovic - Statistique et informatique appliquées, 2e édition, Dunod, Paris, 1973.

79. Levin, Jack - Elementary statistics in social research, New York, Harper and Row, 1973, 279 p.

80. Levinson, Horace C. - Chance, Luck and Statistics, Dover Publications, Inc., New York, 1963.

81. Lewis, Edward Erwin - Methods of statistical analysis in economics and business, Boston, Houghton Mifflin, 1963, 726 p.

82. LI, Jerome C.R. - Statistical Inference, 2 vol., Edwards Brothers, Inc., Ann Arbor, 1969.